

# Hyperanalytic Denoising

Sofia C. Olhede

Department of Mathematics, Imperial College London, SW7 2AZ UK

## Abstract

A new thresholding strategy for the estimation of a deterministic image immersed in noise is introduced. The threshold is combined with a wavelet decomposition, where the wavelet coefficient of the image at any fixed value of the decomposition index is estimated, via thresholding the observed coefficient depending on the value of both the magnitude of the observed coefficient as well as the magnitudes of coefficients of a set of additional images calculated from the observed image. The additional set of images is chosen so that the wavelet transforms of the full set of images have suitable deterministic and joint stochastic properties at a fixed scale and position index. Two different sets of additional images are suggested. The behaviour of the threshold criterion for a purely noisy image is investigated and a universal threshold is determined. The properties of the threshold for some typical deterministic signal structures are also given. The risk of an individual coefficient is determined, and calculated explicitly when the universal threshold is used, and some typical deterministic signal structures. The method is implemented on several examples and the theoretical risk reductions substantiated.

## Index Terms

Image denoising, wavelets, Hilbert transform, 2-D analytic.

## I. INTRODUCTION

**T**HIS paper treats the problem of estimating an unknown deterministic image immersed in noise. The proposed estimation procedure will be based on a separable wavelet decomposition of the observed image that is augmented by a set of wavelet decompositions of additional images calculated from the observed image. The wavelet coefficients of the full set of images at any fixed value of the scale and position are used to estimate the wavelet transform coefficient of the deterministic image at the given scale and position. The transform is then inverted and the spatial domain image estimated. In 1-D signal estimation Donoho and Johnstone [1], [2] first proposed estimation of a noisy deterministic signal using the wavelet transform. The success of such decomposition based methods mainly relies on the deterministic and stochastic properties of the observed or noisy decomposition coefficients at any fixed index value. In the simplest form the estimation procedure roughly corresponds to separating ‘clean’ and ‘noisy’ coefficients into two subsets, where the ‘noisy’ coefficients are eliminated or subjected to some form of shrinkage [3]. Often each coefficient is estimated separately at any given index value and for example the procedure may correspond to eliminating the coefficients whose magnitudes do not exceed a given threshold. A possible choice of threshold is the universal threshold, constant across coefficients, that for large sample sizes gives a risk close to that given by using an ‘oracle,’ *i.e.* knowing whether a coefficient should be eliminated or retained [1]. A slightly different definition is given to the universal threshold by the authors of [4], that we shall use. If the decomposition is highly compressed hard thresholding combined with the universal threshold will achieve very good estimation in terms of low mean square error.

Naturally, to achieve optimal compression for locally simpler structures, such as 1-D behaviour embedded in a 2-D image, whilst still being able to represent varied signal structure in 2-D, the decomposition algorithm becomes more complicated. If a very simple decomposition method is used, then determining the statistical properties of the observed coefficients is easily done, and the decomposition can be found without major computational expense. The draw-back is that in general the mean square error of the estimation will, unless the estimation procedure is more complicated, with a very simply decomposition often increase due to lack of compression. Hence a trade-off must be found between the choice of decomposition and the appropriate treatment of the coefficients of the observed signal to form estimates. This compromise will naturally vary with the assumptions placed on the observed signal.

This work was supported by an EPSRC grant.

S. Olhede is with the Department of Mathematics, Imperial College London, SW7 2AZ, London, UK (s.olhede@imperial.ac.uk). Tel: +44 (0) 20 7594 8568, Fax: +44 (0) 20 7594 8517.

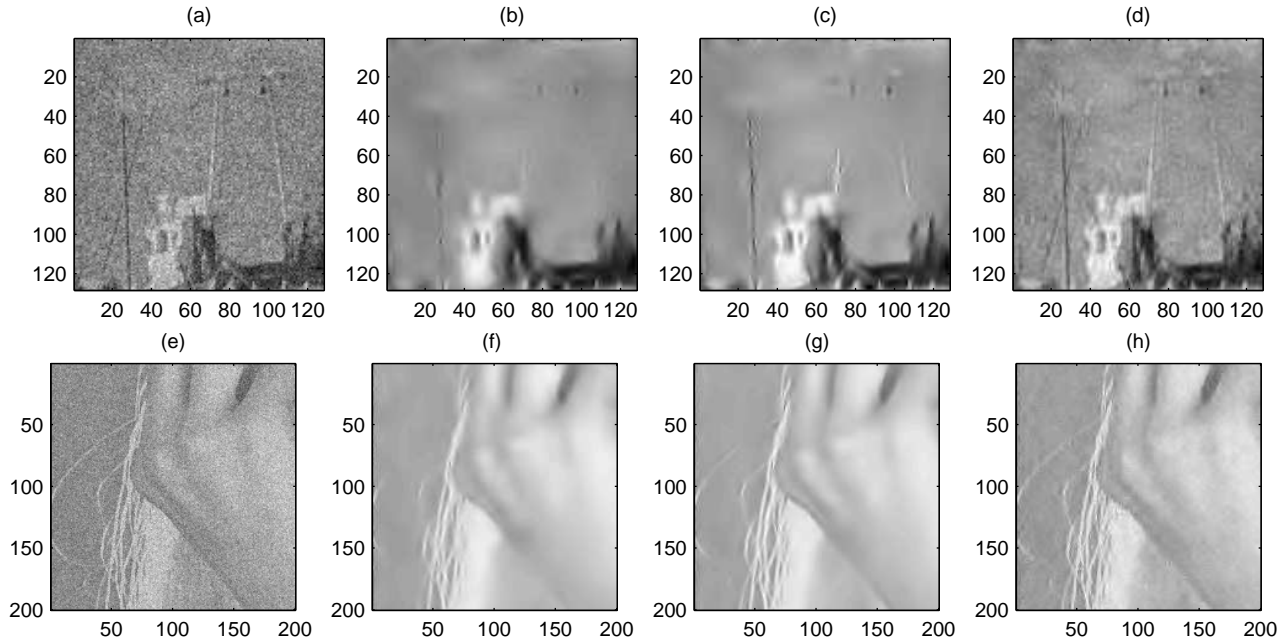


Fig. 1. Row 1: estimating a section of the noisy boat image (SNR=5.56), with the noisy image (a), the usual hard thresholding estimate (b), the hypercomplex estimate (c) and the HMM estimate (d). The hypercomplex estimate in (c) achieves greater continuity, see for example the mast, than that achieved by (b). The HMM estimate has more noise left in the reconstruction, a feature also observed by Starck *et al.* [5][p. 680]. Row 2: a section of one of the bands of the noisy Tiffany image (SNR=8), with the noisy image (a), the usual hard thresholding estimate (b), the hypercomplex estimate (c) and the HMM estimate (d). Observe the curved loose strand of hair, and that the HMM method reconstructs the background with some noisy artifacts.

As the variational structure in 2-D is a great deal richer than in 1-D, many different methods have been developed to achieve optimal compression for given image structures, and for particularly successful examples see work by Starck *et al.* on curvelets [5], work by Donoho on wedgelets [6] or work by le Pennec and Mallat on bandelets [7]. An important feature of all these decompositions is the representation of an image in terms of coefficients associated with a given spatial position, and length scale. The coefficients are considered ‘local’ to such positions and scales.

Methods that achieve a substantial degree of compression can afford to treat each decomposition coefficient individually and without a great deal of sophistication. To achieve better estimation of coefficients for a simplistic decomposition method treating decomposition coefficients *simultaneously* may also give improved estimation. This may correspond to full likelihood based or similar methods such as those proposed by Jansen and Bultheel [8], note also work by Johnstone and Silverman where the local sparsity of decomposition coefficients is discussed [9], as well as usage of specific known coefficient structure in the decomposition of a deterministic signal across coefficients: see for example Cai and Silverman [10], Dragotti and Vetterli [11], Pižurica *et al.* [12], Crouse *et al.* [13], Fryzlewicz [14] and Olhede and Walden [15]. By modelling continuity across coefficients in terms of their local index, estimates with a reduced mean square error may be obtained, that frequently correspond to better visual reconstructions of the image. If the method of estimating the decomposition coefficients is not very complicated but still captures continuity across the decomposition index well, then we may choose a decomposition of the image that is not optimal in terms of compression of the deterministic image, but that is computationally cheap to implement, and where the estimation of the decomposition coefficients may be treated carefully. We may then achieve a reduced mean square error in the estimation of the image at a low computational expense. The purpose of this paper will *not* be to develop an optimal decomposition algorithm, but instead to improve the estimation of the decomposition coefficients, without complicating the procedure substantially. We shall base our image estimate on the 2-D separable wavelet transform coefficients, extending 1-D methods of utilizing coefficient structure to 2-D. We shall make the developments precise by discussing the risk of a given estimated decomposition coefficient.

In 1-D Dragotti and Vetterli [11] explicitly model signal structure as a polynomial function plus some discontinuities and jointly estimate the full set of coefficients corresponding to a discontinuity. Pižurica *et al.* pool

information regarding joint structure in 2-D across coefficients via estimating the local Lipschitz constant and this information is used to estimate the probability that a wavelet coefficient contains contributions from a signal. Crouse *et al.* [13][p. 887] model signal presence across coefficients in 1-D in terms of clustering and persistence, i.e. if a coefficient is non-null at a given decomposition index, then coefficients that are “close” to this index are also non-null. A similar strategy is adopted in 1-D by Cai and Silverman [10], whilst Fryzlewicz [14] considers the magnitude of any additional arbitrary coefficient when estimating a given coefficient. Fryzlewicz established the risk of this strategy, determined from the mean and covariance matrix of the two coefficients. Fryzlewicz’s treatment is very general and instructive.

In a similar spirit to some of the aforementioned methods in 1-D Olhede and Walden [15] considered the thresholding of an individual wavelet coefficient based on the magnitude of the observed coefficient and the magnitude of the decomposition coefficient of the Hilbert Transform (HT) of the signal, denoting this method ‘analytic’ denoising. As both the HT and the wavelet transform are linear the strategy can be viewed either as constructing a second out-of-phase replication of the original signal and finding its local decomposition, or as forming a weighted average of coefficients of the same scale that are nearby in time and using this magnitude to determine if there is local signal presence. The latter strategy is similar in spirit to block thresholding, but instead of using a local magnitude calculated from an average of squared adjacent wavelet coefficients at the same scale in the thresholding procedure, the square of a weighted average of coefficients with a weighting of  $O(1/(2^j k))$  is used. For ‘analytic’ thresholding to work well the wavelet coefficient of the HT of the deterministic signal must be large when the wavelet transform of the signal should not be estimated by zero, and the distribution of the wavelet coefficient of the HT of the noise needs to be jointly determined with the wavelet coefficient of the noise at the same scale and time. As the HT can be considered to have the same time-frequency structure as the original signal the wavelet transform of the signal and the HT of the signal should be large concurrently. Olhede and Walden [15] determined that the wavelet transform of the noise and its HT were approximately uncorrelated at a fixed time and scale, and supplied an appropriate universal threshold for ‘analytic’ denoising. Figure 2(a) shows the risk of a given coefficient using ‘analytic’ denoising, based on the wavelet transform of the signal and its HT taking the same magnitude. The figure verifies that in the case of equivalent means the risk of an ‘analytic’ hard thresholded coefficient estimate with a universal threshold is less than that of a hard thresholded estimate with a universal threshold. Thus improvements to standard denoising in 1-D can be obtained by implementing this procedure.

We seek to extend ‘analytic’ thresholding to 2-D, and this will in general require defining additional images, serving the same purpose as the HT did in 1-D. The HT was useful in 1-D, as it has the same time-frequency structure as the original signal, something we discuss in section IIB, and also the joint statistical properties of the wavelet transform of noise and its HT at a fixed scale and position was easily determined. In 2-D there are many possible extensions to the HT, where in each case more than a single additional component is defined. We refer to such components as *quadrature components*, that are introduced and discussed in sections IIB and IIC. There is more than one extension because variation in the image can either be locally uni-directional, and associated with a given direction, or occurring in several directions simultaneously (see Olhede and Metikas [16] for a more complete discussion of this topic). We investigate the usage of two possible HTs: the Riesz Transforms (RTs, section IID) of the image or the tensor products of the HT in 1-D with the identity filter, denoted the HyperComplex transforms (HCTs, section IIE). We define the local magnitude of the wavelet coefficients from the quadrature components (section IIIA), and propose a threshold criterion to estimate the wavelet coefficients of the image. Once the wavelet transform is inverted this yields an estimate of the image, and this method is denoted hyperanalytic denoising.

We discuss the properties of the local magnitude for stylized image structure: *i.e.* the behaviour of the threshold criterion for oscillatory structures and edges (section IIIB). We discuss the choice of threshold, and an appropriate universal threshold for correlated threshold criteria (section IIIC). We determine the approximate distribution of the decomposition of the Riesz and Hypercomplex components of noise alone at a fixed value of the indexing (section IIID), and this allows us to determine universal thresholds for both the RT and HCT based methods (section IIIE). We calculate the approximate risk associated with the two different thresholding strategies with a given threshold (section IIIF), and discuss the value of the risk of the different procedures for certain scenarios.

We implement the procedure on several examples (section IV), and compare results with the Hidden Markov Model method (HMM). We observe that a reduced mean square error is obtained from using the proposed image denoising strategies, and discernable improvements in the visual reconstructions. Hyperanalytic denoising is thus shown to give a simple and computationally competitive method of improving existing denoising strategies.

## II. IMAGE MODEL

### A. Image Structure

We model the observed image  $[Y_{\mathbf{x}}]_{\mathbf{x}}$  for  $\mathbf{x} = [x_1, x_2]^T \in D$ , where  $D = [0, N-1]^2$ , and  $\Delta x$  denotes the sampling period via:

$$Y_{\mathbf{x}} = q(x_1 \Delta x, x_2 \Delta x) + \epsilon_{\mathbf{x}}, \quad \mathbf{x} \in D. \quad (1)$$

We collect the observed image in a matrix  $\mathbf{Y} = [Y_{\mathbf{x}}]_{\mathbf{x} \in D}$ , and similarly define  $\mathbf{q} = [q_{\mathbf{x}}]_{\mathbf{x} \in D} = [q(x_1 \Delta x, x_2 \Delta x)]_{\mathbf{x} \in D}^T$ , as well as  $\boldsymbol{\epsilon} = [\epsilon_{\mathbf{x}}]_{\mathbf{x} \in D}$ . The noise is modelled by  $\epsilon_{\mathbf{x}} \sim N(0, \sigma^2)$ , where  $\sim$  denotes distributed as, and  $\text{Cov}(\epsilon_{\mathbf{x}}, \epsilon_{\mathbf{y}}) = \sigma^2 \delta_{\mathbf{x}, \mathbf{y}}$ ,  $\mathbf{x}, \mathbf{y} \in D$ , i.e. the noise is Gaussian, uncorrelated and isotropic. A decomposition of the image in terms of a wavelet basis [17] is formed via

$$q(\mathbf{x} \Delta x) = \sum_{j, \mathbf{k}} W_{j,1,\mathbf{k}}^{(q)} \psi_{j,1,\mathbf{k}}(\mathbf{x}) + \sum_{j, \mathbf{k}} W_{j,2,\mathbf{k}}^{(q)} \psi_{j,2,\mathbf{k}}(\mathbf{x}) + \sum_{j, \mathbf{k}} W_{j,3,\mathbf{k}}^{(q)} \psi_{j,3,\mathbf{k}}(\mathbf{x}) + \sum_{\mathbf{k}} V_{k_1, k_2}^{(q)} \phi_{J,\mathbf{k}}(\mathbf{x}), \quad (2)$$

where  $\psi_{j,1,\mathbf{k}}(\mathbf{x})$ ,  $\psi_{j,2,\mathbf{k}}(\mathbf{x})$ ,  $\psi_{j,3,\mathbf{k}}(\mathbf{x})$ , and  $\phi_{j,\mathbf{k}}(\mathbf{x})$  are the tensor products of functions  $\psi_{j,k}(x)$  and  $\phi_{j,k}(x)$ , respectively.  $V_{J,\mathbf{k}}^{(q)} \equiv W_{J,4,\mathbf{k}}^{(q)}$  is then associated with smooth behaviour in the image  $q(\mathbf{x})$  in the variables  $x_1$  and  $x_2$ ,  $W_{j,2,\mathbf{k}}^{(q)}$  is associated with smooth behaviour in  $x_2$  and rapid variation in  $x_1$ , *etc*, where  $u = 1, \dots, 4$  denotes the tensor product index.  $j$  is associated with scale  $2^j$ , where  $1 \leq j \leq J_0 = \lg(N)$ , whilst  $\mathbf{k}$  is associated with a spatial localisation in the plane. If an image with  $N^2$  coefficients is observed, then for any fixed value  $j$ ,  $0 \leq k_l \leq N_j - 1$ ,  $l = 1, 2$ , where  $N_j = N/2^j$ . For simplicity collect the indices in a vector-valued index of  $\boldsymbol{\xi} = [j, u, \mathbf{k}]^T$ . The full set of coefficients  $[W_{j,u,\mathbf{k}}^{(q)}]$  is the Discrete Wavelet Transform (DWT) of  $q(\cdot)$ .

The DWT is usually implemented by repeated filtering of the observed signal with two special filters, the scaling filter  $\{g_l : l = 0, \dots, L-1\}$  and the wavelet filter  $\{h_l : l = 0, \dots, L-1\}$ , in both spatial directions separately. We initialise the transform by equating the image with the finest scale representation of the image, i.e.  $V_{0,x_1,x_2}^{(q)} \equiv q_{x_1,x_2}$ . The transform at index  $\boldsymbol{\xi}$  can also be implemented using a single filter  $h_{j,l}$ . The decomposition is halted at level  $j = J \leq J_0 = \lg(N)$ , and the scaling coefficients  $\{V_{J,\mathbf{k}}^{(q)}\}$  are determined at this level to complete the representation. Hence for  $j < J$  only  $[W_{j,u,\mathbf{k}}]$  for  $u = 1, 2, 3$  are calculated. For more details on the DWT, see for example Percival & Walden [18], whilst a good exposition of image decompositions can be found in Mallat [17]. Having observed  $Y_{\mathbf{x}}$  rather than  $q_{\mathbf{x}}$  we calculate the DWT coefficients  $W_{\boldsymbol{\xi}}^{(Y)}$ , and threshold these to obtain an estimate of  $W_{\boldsymbol{\xi}}^{(q)}$ , denoted  $\widehat{W}_{\boldsymbol{\xi}}^{(q)}$ . Wavelets will compress images of sufficient regularity, a statement that can be made precise in terms of Besov spaces, but for some locally simple image structures, a more compressed representation can be made [5], [7]. Hence for images containing say edges the deterministic image energy in the DWT will be spread over more coefficients than strictly necessary, and as the magnitude of the affected coefficients will be less than the coefficients representing the same structure in a more compressed alternative decomposition it is important that the estimation procedure does not fail to retain signal generated coefficients.

### B. Quadrature Components

In one version of the 1-D estimation algorithms suggested by Cai and Silverman [10], the coefficient at scale  $j$  and position  $k$  was estimated using a shrinkage rule depending on the combined magnitude of the observed coefficient at  $[j, k]$  and the magnitude of the immediate time-neighbours at the same scale, i.e. at  $[j, l]$  for  $l \neq k$ . This procedure will perform well if a signal contribution present at the  $[j, k]$  index exhibits clustering in adjacent coefficients, i.e. the wavelet coefficients will have large means at  $[j, l]$ , and the noise is uncorrelated over  $l \neq k$ . The scale adjacent coefficients at a given time point have been used to improve estimation [13], [14], and Olhede and Walden [15] used the wavelet decomposition of the HT of the observed image to this purpose. We seek to generalise the method in [15] to 2-D, and discuss some of its properties, before proceeding to do so.

To simplify the discussion of the HT, let the Fourier Transform (FT) of a  $d$  dimension signal  $q(\mathbf{x})$  be denoted by:

$$Q(\mathbf{f}) = \int_{\mathbb{R}^d} q(\mathbf{x}) e^{2i\pi \mathbf{f}^T \mathbf{x}} d^d \mathbf{x} = |Q(\mathbf{f})| e^{-2i\pi \varphi_q(\mathbf{f})},$$

this defining the magnitude ( $|Q(\mathbf{f})|$ ) and phase ( $\varphi_q(\mathbf{f})$ ) of  $q(\mathbf{x})$  in the Fourier domain. Given a 1-D signal  $q(x)$  the HT in both the time and frequency domain are defined by:

$$\mathcal{H}q(x) = \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{q(y)}{x-y} dy, \quad (\mathcal{H}Q)(f) = (-i)Q(f) \operatorname{sgn}(f), \quad (3)$$

and the transform can be approximated suitably for discrete implementation (see [15]). Usually  $q(x)$  and  $\mathcal{H}q(x)$  are collected into a complex-valued representation, denoted the analytic signal, given by  $q^{(+)}(x) = q(x) + i\mathcal{H}q(x)$ . If  $q(x) = \cos(2\pi f'x)$  then  $q^{(+)}(x) = \exp(2\pi i f'x)$ , but sometimes too much emphasis is put on this description of oscillatory signals, to the extent where the HT is almost discounted in usage when the observed signal is not oscillatory. Even if  $q(x)$  does *not* correspond to an oscillation, the HT can be considered to enjoy certain properties, such as: i)  $\mathcal{H}q(x)$  is orthogonal to  $q(x)$ , i.e.  $\int \mathcal{H}q^*(x)q(x) dx = 0$ , ii) the magnitude of the HT of  $q(x)$  at any given frequency  $f \neq 0$  is identical to that of the original signal, i.e.  $|Q(f)|^2 = |\mathcal{H}Q(f)|^2$ , iii) the HT is linear in the signal, and iv) the HT of a signal can be considered as having the same time-frequency signature as the original signal. i-iii) immediately follow from equation (3), and ensure that the distribution of the DWT of the HT of white noise at a given value of  $[j, k]$  is asymptotically identical to that of the DWT of the original noise, and the two wavelet coefficients are approximately uncorrelated [15]. The fourth property merits some further discussion. Clearly the notion that a signal and its HT have the same time-frequency structure is accepted in signal processing, as the analytic signal, rather than the real signal, is used to construct time-frequency representations of a real signal. For example usage of the Wigner-Ville distribution rather than the Wigner distribution is generally advocated [19]. As may be noted from equation (3)  $\mathcal{H}(q)(x)$  has at all frequencies  $f \neq 0$  exactly the same frequency support as  $q(x)$ , whilst the spatial support of  $q(x)$ , has been spread out by the convolution with  $1/(\pi x)$ .

The HT is usually interpreted as a phase-shift of  $\pi/2$  to signal  $q(x)$ . Note that we may write:

$$q(x) = 2 \int_0^{\infty} |Q(f)| \cos(2\pi(fx - \varphi_q(f))) df, \quad \mathcal{H}q(x) = 2 \int_0^{\infty} |Q(f)| \sin(2\pi(fx - \varphi_q(f))) df. \quad (4)$$

Hence the same magnitude of  $|Q(f)|$  is assigned to each frequency  $f$ , and the contribution previously associated with  $\cos(2\pi(\varphi_q(f) - fx))$  is now shifted in cycle or phase by  $\pi/2$ . Thus in some sense, we are recovering the same signal, as the frequency description is the same, but there has been a very slight shift in time alignment. Thus  $\mathcal{H}q(x)$  should have roughly the same time-frequency support as  $q(x)$ . This implies that the DWT coefficient of  $q(x)$  should have about the same magnitude as  $\mathcal{H}q(x)$ , as the DWT forms a time-frequency decomposition of a given signal. The DWT is compact in time, and we wish to encourage using time information in nearby locations when estimating a given coefficient in analogue with Cai & Silverman. Of course once the coefficient has been estimated, the estimate of the signal will be based on the thresholded wavelet coefficients of the observed signal alone, and thus discontinuities can still be reconstructed.

Given the nice deterministic and stochastic properties of ‘analytic’ denoising, it is not strange that we seek to generalise the concept to 2-D. A first step in this procedure is the definition of linear transformations of the image that will serve the same purpose as the HT did in 1-D. The HT and the signal formed a natural representation in terms of the ‘analytic’ signal, where the real and imaginary components were phase shifted versions of each other, or we may denote the latter two signals as being ‘in quadrature,’ i.e. as representing out-of-phase replications of the same structure. Their magnitude squares represent the local presence of the signal well but we stress that even if the signal is not oscillatory, the interpretation of the HT as having roughly the same time-frequency support still rests on the above arguments. We shall denote the signal and its HT as quadrature components, and will define the 2-D generalization of this two-component signal collection.

**Definition 2.1 (Quadrature Components):** We denote by *Quadrature Components* of  $q(\mathbf{x})$  any set of images  $\{\check{q}^{(s,l)}(\mathbf{x})\}_{l=1}^L$ , where  $s$  denotes the specific transform used in the construction of the components that satisfy:

- 1) each  $\check{q}^{(s,l)}(\mathbf{x})$  is orthogonal (‘out of phase’) to the original signal  $q(\mathbf{x}) \equiv \check{q}^{(s,0)}(\mathbf{x})$ , or  $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} q^*(\mathbf{x}) \check{q}^{(s,l)}(\mathbf{x}) d^2\mathbf{x} = 0$ , and for all separable  $q_S(\mathbf{x}) = q_1(x_1)x_2(x_2)$ , also  $\int_{-\infty}^{\infty} q_S^*(\mathbf{x}) \check{q}_S^{(s,l)}(\mathbf{x}) dx_l = 0$ ,
- 2) the combined energy assigned to each frequency  $\mathbf{f}$  from the full set of quadrature components at all points of  $\mathbf{f} \in \mathbb{R}^2$  except for a finite set of frequencies satisfies the equation  $\sum_{l=1}^L \left| \check{Q}^{(s,l)}(\mathbf{f}) \right|^2 = C_L^{(s)} |Q(\mathbf{f})|^2$ , where  $0 < C_L^{(s)} < \infty$  is constant and,

- 3) each  $\check{q}^{(s,l)}(\mathbf{x})$ , for  $l = 1, \dots, L$ , is constructed by a linear transformation of  $q(\mathbf{x})$ ,  
 4) the space and spatial frequency support of  $[\check{q}^{(s,1)}(\mathbf{x}), \dots, \check{q}^{(s,L)}(\mathbf{x})]$ , for  $l = 1, \dots, L$  is similar to that of  $q(\mathbf{x})$ .  
 We form the DWT of all  $L + 1$  images,  $\left\{ \check{q}_{\mathbf{x}}^{(s,l)} \right\}_{l=0}^L$ , and define:

$$\check{W}_{\xi}^{(q,s,l)} = W_{\xi}^{(\check{q}^{(s,l)})}, \quad l = 0, \dots, L. \quad (5)$$

The linear operator that constructs object  $\check{q}^{(s,l)}(\mathbf{x})$  from  $q(\mathbf{x})$  will be denoted  $\mathcal{V}^{(s,l)}$  and the transformation is implemented in the spatial domain using the kernel  $v^{(s,l)}(\mathbf{x})$  that once the integral is approximated using a Riemann sum is replaced by a linear filter  $v_{D,\mathbf{x}}^{(s,l)}$ . The FT of  $v^{(s,l)}(\mathbf{x})$  is denoted  $V^{(s,l)}(\mathbf{f})$  whilst taking the FT of  $v_{D,\mathbf{x}}^{(s,l)}$  yields the object  $V_D^{(s,l)}(\mathbf{f})$ . The discrete implementation of the calculation of the quadrature components is outlined in Appendix A.

### C. Stochastic Properties of Decomposed Quadrature Components

We establish the stochastic properties of the wavelet decomposition of noise alone, and for this purpose define at a fixed value of  $\xi$ :

$$\mathbf{n}^{(s,u)} = \left[ \check{W}_{\xi}^{(\epsilon,s,0)}, \dots, \check{W}_{\xi}^{(\epsilon,s,L)} \right]^T, \quad s = r, h. \quad (6)$$

*Proposition 1 (Energy of Quadrature Components):* At a fixed index value  $\xi = [j, u, \mathbf{k}]^T$  the total energy of the DWT of the quadrature components of white noise with variance  $\sigma^2$  is given by:

$$\mathbb{E} \left( \sum_{l=1}^L n_l^{(s,u)2} \right) = C_L^{(s)} \sigma^2 + O(1/N), \quad u = 1, 2, 3, 4. \quad (7)$$

*Proof:* See appendix C. The error term follows from the Riemann approximation to the integral. ■

*Proposition 2 (Covariance of Transforms of the Signal and Its Quadrature Components):* At a fixed index value  $\xi = [j, u, \mathbf{k}]^T$  the covariance of the DWT of white noise, and the DWT of any of the quadrature components of the white noise is of order  $O(1/N)$ .

*Proof:* See appendix C. The error term follows from the Riemann approximation to the integral. ■

Thus at any given value of  $\xi$  the DWTs of  $\epsilon_{\mathbf{x}}$  and  $\left\{ \check{\epsilon}_{\mathbf{x}}^{(s,l)} \right\}$  are approximately uncorrelated, and the combined energy of the DWTs of  $\left\{ \check{\epsilon}_{\mathbf{x}}^{(s,l)} \right\}$  is a multiplicative constant of the energy of the DWT of  $\epsilon_{\mathbf{x}}$ . Thus the squared magnitudes of these objects have a tractable joint distribution. Condition 4 ensures that we may assume that the mean of the DWT of the observed image will be simultaneously large to the mean of the DWTs of the quadrature components of the observed image at a given value of the index  $\xi$ . Of course whilst the general definition of ‘quadrature components’ may then seem justifiable, this does not guarantee the existence of such objects. We shall give two different specific examples of quadrature components based on extending the HT to 2-D, and discuss their properties. We base the set of quadrature components on hyperanalytic functions, see [16].

### D. The Riesz Transforms

The Riesz Transforms (RTs) have been used in combination with the wavelet transform by Metikas and Olhede [16], [20]. Denote the convolution of two functions  $q_1(\mathbf{x})$  and  $q_2(\mathbf{x})$  by  $(q_1 * q_2)(\mathbf{x}) = \int \int_{\mathbb{R}^2} q_1(\mathbf{y}) q_2(\mathbf{x} - \mathbf{y}) d^2 \mathbf{y}$ . The RTs of  $q(\cdot)$ , denoted  $\check{q}^{(r,1)}(\mathbf{x})$  and  $\check{q}^{(r,2)}(\mathbf{x})$ , are obtained by convolving  $q(\cdot)$  with the Riesz kernels  $v^{(r,l)}(\mathbf{x})$ , given in terms of  $x = \sqrt{x_1^2 + x_2^2}$  and  $f = \sqrt{f_1^2 + f_2^2}$  by:

$$v^{(r,l)}(\mathbf{x}) = \frac{x_l}{2\pi x^3}, \quad V^{(r,l)}(\mathbf{f}) = -i \frac{f_l}{f}, \quad l = 1, 2, \quad \check{q}^{(r,l)}(\mathbf{x}) = \left( v^{(r,l)} * q \right)(\mathbf{x}), \quad l = 1, 2. \quad (8)$$

The RTs satisfy the conditions of quadrature components, see for example [16][p. 15–16]. Given the RTs combine to have the same norm as  $q(\mathbf{x})$ ,  $C_2^{(r)} = 1$ . As in Olhede & Metikas [16] we argue that if unidirectional structure only is present in the image with orientation  $\nu$ , i.e. the image admits the representation for  $\mathbf{n} = [\cos(\nu) \quad \sin(\nu)]^T$  and  $\nu \in (0, \pi)$  of:

$$q_U(\mathbf{x}) = \int_0^\infty G_U(f) \cos(2\pi \mathbf{f}^T \mathbf{n} \mathbf{x}) df, \quad (9)$$

then the interpretation of the RTs is simplified. We use polar coordinates and set  $\mathbf{f} = f [\cos(\phi) \quad \sin(\phi)]$ . Then the Fourier transform of  $q_U(\mathbf{x})$  is  $Q_U(\mathbf{f}) = \frac{G_U(f)}{2} [\delta(\phi - \nu) + \delta(\phi + \nu)]$  and we find:

$$\left[ \check{q}_U^{(r,1)}(\mathbf{x}) \check{q}_U^{(r,2)}(\mathbf{x}) \right] = [\cos(\nu) \quad \sin(\nu)] \int_0^\infty G_U(f) \sin(2\pi f \mathbf{n}^T \mathbf{x}) df. \quad (10)$$

Thus the two quadrature components represent the same 1-D directional variation as  $q_U(\mathbf{x})$ , with the same directionality as  $q_U(\mathbf{x},)$  but where the variations in direction  $\nu$  have been shifted in phase and multiplied by a constant factor. Thus (informally) for unidirectional variation the Riesz transforms have the same spatial and spatial frequency support as the original signal. Note that we are *not* assuming that  $q(\mathbf{x})$  is periodic or oscillatory.

### E. The Hypercomplex Transforms

A second set of 2-D HTs are the HyperComplex Transforms (HCTs), defined as tensor products of the identity transform and the HTs. By Olhede and Metikas [16][p. 12–13], it is shown that the hypercomplex transforms give a valid set of quadrature components, and note  $C_3^{(h)} = 3$ . Denote the partial HT [21] in direction  $x_l$  by  $\mathcal{H}_l$ . Three additional quadrature components are defined by:

$$\check{q}^{(h,1)}(\mathbf{x}) = \mathcal{H}_1 \{q\}(\mathbf{x}), \quad \check{q}^{(h,2)}(\mathbf{x}) = \mathcal{H}_2 \{q\}(\mathbf{x}), \quad \check{q}^{(h,3)}(\mathbf{x}) = \mathcal{H}_2 \mathcal{H}_1 \{q\}(\mathbf{x}). \quad (11)$$

If the image is naturally expressed as separable in the frame of reference the three HCTs of  $q_S(\cdot)$  are by trivial extension of equation (4), the same signal shifted in phase in the two axes. Of course the purpose of this paper will be to alleviate problems (see for example Starck *et al.* [5][p. 671]) when estimating nonseparable images based on coefficients calculated in a separable decomposition whose energy spread across more coefficients than strictly necessary. Assume  $q(\mathbf{x})$  is non-separable then define its Partial FT (PFT) in direction  $x_1$  by:  $Q_1(f_1, x_2) = |Q_1(f_1, x_2)| e^{-2\pi i \varphi_1(f_1, x_2)} = \int_{-\infty}^\infty q(\mathbf{x}) e^{-2\pi i f_1 x_1} dx_1$ , so

$$q(\mathbf{x}) = 2 \int_0^\infty |Q_1(f_1, x_2)| \cos(2\pi(f_1 x_1 - \varphi_1(f_1, x_2))) df_1 \quad (12)$$

$$\mathcal{H}_1 q(\mathbf{x}) = 2 \int_0^\infty |Q_1(f_1, x_2)| \cos(2\pi(f_1 x_1 - \varphi_1(f_1, x_2) - \pi/2)) df_1. \quad (13)$$

Thus  $\check{q}^{(h,1)}(\mathbf{x})$  corresponds to replicating all variation in  $x_1$ , for any fixed value of  $x_2$ , but shifted in phase, and *mutatis mutandis* the analogous statements hold for  $\check{q}^{(h,2)}(\mathbf{x})$  and  $\check{q}^{(h,3)}(\mathbf{x})$ . If  $q(\mathbf{x})$  corresponds to a particular time-frequency structure as a signal in  $x_1$  for fixed values of  $x_2$  then  $\check{q}^{(h,1)}(\mathbf{x})$  will replicate the same structure, but shifted in phase, in analogue with equation (4). We propose to use the decomposition coefficients of the three HCTs of the observed image to estimate the decomposition of the deterministic image. An improvement in estimation will ensue if the magnitudes of the decomposition coefficients of the HCTs are large when the coefficient of the observed image should be kept rather than killed. Given each coefficient replicates the same variational structure in each separate axes this should be the case, cf equation (4). For simple 1-D structures such as line segments observed in 2-D the energy of the image will be spread over more coefficients than strictly necessary. By defining the additional images that should have the same marginal variational structure as the images in each of the two axes, for moderate SNRs the estimation should improve by using the additional components, as more often the signal is recognized as present. Subsequent risk calculations in section III-F show that if the DWT of the quadrature components have the same mean as the DWT of the original signal, then the risk can be reduced by a new procedure that uses the magnitudes of all four components at each fixed  $\xi$ , to threshold a given coefficient. The authors of [15] denoised a 1-D signal by defining a second component as the HT of the observed signals, and using the DWT of this component when thresholding the observed signal. A naive 2-D extension of this method would define a single extra quadrature component corresponding to  $\check{q}^{(h,3)}(\mathbf{x})$ , phase shifting in both spatial directions simultaneously. We discuss the risk of this procedure in section III-F, and it is shown to exceed that of the proposed method, for certain scenarios.

### III. ESTIMATION

#### A. Defining Estimates

We have argued that the quadrature components defined either by the HCTs or RTs have the same space and spatial frequency structure as the original image, shifted in phase. Therefore the mean of the DWTs of the quadrature components should be the same as the DWT of the signal. We define a local magnitude in terms of the DWTs of the full set of quadrature components.

*Definition 3.1 (The Magnitude of a Coefficient):* We define the magnitude of a coefficient  $W_{\xi}^{(q)}$  using quadrature components denoted by  $s$  via:  $M_{\xi}^{(q,s)2} = \frac{1}{C_L^{(s)}+1} \sum_{l=0}^L W_{\xi}^{(q,s,l)2}$ .

Let for some fixed  $L_1 \in \mathbb{N}$   $B_i^j = \{k : k = i + l, l = -L_1, \dots, L_1\}$ , then  $M_{\xi}^{(q,s)2}$  is a 2-D analogy to  $S_{j,k}^2 = \sum_{l \in B_k^j} W_{[j,l]}^{(q)2}$ , used by Cai and Silverman [10][p. 132], to block threshold. Each coefficient  $W_{\xi}^{(q)}$  will be estimated by hard thresholding the observed coefficient depending on the value of  $M_{\xi}^{(Y,s)2}$

$$\widehat{W}_{\xi}^{(q,s)}(\lambda^2) = \begin{cases} W_{\xi}^{(Y)} & \text{if } M_{\xi}^{(Y,s)2} \geq \frac{\sigma^2}{C_L^{(s)}+1} \lambda^2 \\ 0 & \text{if } M_{\xi}^{(Y,s)2} < \frac{\sigma^2}{C_L^{(s)}+1} \lambda^2 \end{cases} \quad (14)$$

The notation given  $\xi$  is fixed when estimating any set coefficient is needlessly complicated, and we remove the reference to most of these indices. We define:  $\mathbf{W}^{(s,u)} = [W_{\xi}^{(Y,s,0)}, \dots, W_{\xi}^{(Y,s,L)}]^T$ ,  $\boldsymbol{\mu}^{(s,u)} = [W_{\xi}^{(q,s,0)}, \dots, W_{\xi}^{(q,s,L)}]^T$ , note that  $\mu_1 = \mu_1^{(s,u)}$  does not depend on the choice of  $s$  and we denote the estimator using the  $s$  indexed components by:  $\widehat{\mu}_1^{(s,u)}(\lambda^2) = \widehat{W}_{\xi}^{(q,s)}(\lambda^2)$ .

#### B. Magnitude of Typical Image Features

Deterministic images are frequently modelled as the combination of texture and contours (see for example work by Vese and Osher [22] modelling a signal as a bounded variation contribution plus a texture contribution). We consider observing an image that is an aggregation of edges and texture, where each texture component is modelled by  $t(\mathbf{x}) = a_t(\mathbf{x}_0) \cos(2\pi \mathbf{f}_0^T \mathbf{x})$ , and each edge component is modelled by  $e(\mathbf{x}) = a_e(x_1) \delta(\cos(\theta)x_1 + \sin(\theta)x_2 - c)$ .  $a_e(\cdot)$  and  $a_t(\cdot)$  are assumed to be slowly varying. In general we do not expect to observe sinusoids or discontinuities that for very slowly varying  $a_e(\cdot)$  span the entire observed image, but to be able to carry out theoretical calculations stylized image structures must be analysed, that observed images would approximate. We shall investigate how the magnitude of the transform coefficients of the full set of quadrature components behave, as our subsequent risk calculations will demonstrate that the success of the method strongly depends on the mean of the quadrature components.

To this purpose we define the Maximum Overlap Discrete Wavelet Transform Coefficients (MODWT coefficients)  $\widetilde{W}_{\xi}^{(e,s,l)}$ . These are the DWT coefficients calculated without subsampling, and where a new normalisation is introduced at each level  $j$  to preserve energy. For a full length discussion see Percival and Walden [18][Ch. 4]. We denote the FT of the MODWT filter  $\tilde{h}_{j,u,\mathbf{k}}$  by  $\tilde{H}_{j,u}(\mathbf{f}) = \left| \tilde{H}_{j,u}(\mathbf{f}) \right| e^{-2i\pi \tilde{\varphi}_{j,u}(\mathbf{f})}$ , thus defining the modulus ( $|\tilde{H}_{j,u}(\mathbf{f})|$ ) and phase ( $\tilde{\varphi}_{j,u}(\mathbf{f})$ ) of  $\tilde{H}_{j,u}(\mathbf{f})$ . For notational convenience let  $\mathbf{x} = 2^j(\mathbf{k} + 1) - 1$ , and the region of frequency space where  $\tilde{H}_{j,u}(\mathbf{f})$  is mainly supported be denoted  $\Omega_{j,u}$ . The DWT coefficients of a generic signal  $q(\mathbf{x})$  can be extracted from the MODWT coefficients of  $q(\mathbf{x})$ , using the relations  $W_{j,u,\mathbf{k}}^{(q)} = 2^j \tilde{W}_{j,u,\mathbf{x}}^{(q)}$  (see for example Percival and Walden [18][p. 203]).

*Lemma 1 (RT Magnitude of Local Oscillation):* If the signal locally takes the form  $t(\mathbf{x}) = a_t(\mathbf{x}_0) \cos(2\pi \mathbf{f}_0^T \mathbf{x})$  with  $\mathbf{f}_0 = [f_0 \cos(\phi_0) \quad f_0 \sin(\phi_0)]^T$ , then the magnitude of the wavelet decomposition defined by definition 3.1 is given by:

$$M_{\xi}^{(t,r)2} = 2^{2j-1} a_t^2(\mathbf{x}_0) I(\mathbf{f}_0 \in \Omega_{j,u}) + \rho_1 + O\left(\frac{1}{N}\right), \quad (15)$$

where  $\rho_1$ , is an error term depending on the leakage of the wavelet filters in the frequency domain. If a sufficiently long wavelet filter is used, this term can be ignored. See Nielsen [23] for more discussion on avoiding leakage.

*Proof:* See appendix B. ■



*Lemma 2 (HCT Magnitude of Local Oscillation ):* If the signal locally takes the form  $t(\mathbf{x}) = a_t(\mathbf{x}_0) \cos(2\pi \mathbf{f}_0^T \mathbf{x})$  with  $\mathbf{f}_0 = [f_0 \cos(\phi_0) \quad f_0 \sin(\phi_0)]^T$ , then the magnitude defined in definition 3.1 is given by:

$$M_{\xi}^{(t,h)2} = 2^{2j-1} a_t^2(\mathbf{x}_0) I(\mathbf{f}_0 \in \Omega_{j,u}) + \rho_2 + O\left(\frac{1}{N}\right), \quad (16)$$

where  $\rho_2$ , is an error term depending on the leakage of the wavelet filters in the frequency domain.

*Proof:* See appendix B. ■

*Lemma 3 (RT Magnitude of Discontinuity ):* If the signal locally can be approximated by  $e(\mathbf{x}) = a_e(x_1) \delta(\cos(\theta)x_1 + \sin(\theta)x_2 - c)$ , then the magnitude defined in definition 3.1 is given by:

$$\begin{aligned} W_{\xi}^{(e)} &= 2^j A_e(0) \int_{\nu_{1,\min}}^{\nu_{1,\max}} \tilde{H}_{j,u}(\cos(\theta)\nu_1, \sin(\theta)\nu_1) e^{2\pi i \nu_1 (\cos(\theta)x_1 + \sin(\theta)x_2 - c)} d\nu_1 + \rho_3 + O\left(\frac{1}{N}\right) \\ U_{\xi}^{(e)} &= 2^j A_e(0)(-i) \int_{\nu_{1,\min}}^{\nu_{1,\max}} \text{sgn}(\nu_1) \tilde{H}_{j,u}(\cos(\theta)\nu_1, \sin(\theta)\nu_1) e^{2\pi i \nu_1 (\cos(\theta)x_1 + \sin(\theta)x_2 - c)} d\nu_1 \\ &\quad + \rho_4 + O\left(\frac{1}{N}\right), \quad M_{\xi}^{(e,r)2} = \frac{1}{2} W_{\xi}^{(e)2} + \frac{1}{2} U_{\xi}^{(e)2} + \rho_5 + O\left(\frac{1}{N}\right) \end{aligned} \quad (17)$$

where  $\rho_3$ ,  $\rho_4$  and  $\rho_5$  depend on the smoothness of  $a_e(\cdot)$ , whilst  $\nu_{1,\min}$  and  $\nu_{1,\max}$  are given in appendix B.

*Proof:* See appendix B. ■

*Lemma 4 (HCT Magnitude of Discontinuity ):* If the signal locally can be approximated as a discontinuity  $e(\mathbf{x}) = a_e(x_1) \delta(\cos(\theta)x_1 + \sin(\theta)x_2 - c)$ , then the magnitude defined in definition 3.1 is given by:

$$M_{\xi}^{(e,r)2} = \frac{1}{2} W_{\xi}^{(e)2} + \frac{1}{2} U_{\xi}^{(e)2} + \rho_6 + O\left(\frac{1}{N}\right), \quad (18)$$

where  $\rho_6$  depends on the smoothness of  $a_e(\cdot)$ , and the forms of  $W_{\xi}^{(e)}$  and  $U_{\xi}^{(e)}$  are given in lemma 3.

*Proof:* See appendix B. ■

For oscillatory signals the magnitude hence aptly reflects signal presence at  $\xi$ . Equation (17) illustrates the problem experienced by an edge in a 2-D separable representation: only if  $\theta = 0$  or  $\theta = \pi/2m$  for  $m \in \mathbb{Z}$  will the edge live in  $\Omega_{j,2}$  or  $\Omega_{j,3}$ , *i.e.* constant in one direction and variable in the other. Only in this case will the representation be extremely compressed (note that the proof needs to be adjusted for  $\theta = 0$ ). From equation (17) we note that the compact spatial support of  $\tilde{h}_{\xi_x}$  ensures that the energy of  $U_{\xi}^{(e)}$  and  $W_{\xi}^{(e)}$  are mainly concentrated near  $\cos(\theta)x_1 + \sin(\theta)x_2 = c$ . Given we may represent  $\tilde{H}_{j,u}(\cdot)$  in terms of a magnitude and a phase, the difference between  $\tilde{h}_{j,u}(\cos(\theta)\nu, \sin(\theta)\nu)$  and  $2\pi\nu(\cos(\theta)x_1 + \sin(\theta)x_2 - c)$  will determine exactly at which spatial indices  $W_{\xi}^{(e)}$  and  $U_{\xi}^{(e)}$  have non-negligible magnitudes (see for example the discussion in Gopinath [24][p. 1794]). Thus the DWT of the quadrature components will be large near the discontinuity and they can be used to improve the estimation. Our proposed procedure will capitalise on this fact, and it can be noticed in the reconstructions that line discontinuities are better reconstructed (see Figure 1 (c)), and as a curved discontinuity can be approximated as the aggregation of amplitude modulated line discontinuities, improvements in estimation can be observed for curved structures (see Figure 1 (g)).

### C. Distribution of Noise & Universal Thresholds

The distribution of  $\mathbf{n}^{(s,u)}$  must be determined, to obtain a universal threshold [4]. Let  $K = N^2$  be the total number of coefficients of the original observed image and denote  $M_K^{(s,\max)} = \max_{\xi} (C_L^{(s)} + 1) M_{\xi}^{(s)2} / \sigma^2$ . Downie and Silverman [4] proposed that a universal threshold should in general satisfy taking a value such that  $\lim_{K \rightarrow \infty} P\left(M_K^{(s,\max)} \leq \lambda_K^2\right) = C$ , for some constant  $0 < C < 1$ , and as  $K$  increases the expected number of coefficients exceeding the threshold is some small but finite non-zero value. We choose a slightly more conservative threshold, so that if our strategy were to be used for  $K$  independent threshold criteria, then the probability that the maximum exceeded the threshold is  $O\left(\frac{1}{(\log(K))^{3/2}}\right)$ , rather than tending to a positive constant for the Riesz threshold. We use for the Hypercomplex threshold a conservative version of that suggested by Downie and Silverman. We cannot quite achieve analogous results to Downie and Silverman as as the set of DWT coefficients  $\mathbf{n}^{(s,u)}$

are correlated across indices  $\xi$ , and adopt arguments similar to those given by Johnstone and Silverman [25] and Olhede & Walden [15], to justify the choice of threshold. We do not aim to determine the full covariance structure of the full set of wavelet coefficients of the observed image and the quadrature components. To derive the conservative threshold define  $\mathcal{M}_K^{(s)}$  as the maximum of  $K$  independent variates with the same marginal distribution as  $\{(C_L^{(s)} + 1)M_\xi^{(s)2}/\sigma^2\}$ . We find a universal threshold based on determining the distribution of  $\mathcal{M}_K^{(s)}$ , and this then constitutes a conservative choice for  $M_K^{(s,\max)} = \max_\xi (C_L^{(s)} + 1)M_\xi^{(s)}/\sigma^2$ ,  $s = r, h$  as

$$\begin{aligned} P\left(M_K^{(s,\max)} \leq \lambda_K^{(s)2}\right) &= P\left(\cap_{i=1}^K \left[\frac{M_\xi^{(s)2}}{\sigma^2/(C_L^{(s)} + 1)} \leq \lambda_K^{(s)2}\right]\right) \geq \prod_{i=1}^K P\left(\frac{M_\xi^{(s)2}}{\sigma^2/(C_L^{(s)} + 1)} \leq \lambda_K^{(s)2}\right) \\ &= P\left(\mathcal{M}_K^{(s)} \leq \lambda_K^{(s)2}\right), \end{aligned}$$

by corollary 2 from Dykstra [26].  $C_L^{(s)} + 1$  can be interpreted as the degrees of freedom associated with  $M_\xi^{(s)2}$ . As shown in the subsequent section, the DWT of the quadrature components for any fixed value of  $\xi$  are uncorrelated, and the  $\mathbf{A}_i$  matrices of Dykstra are defined to take the variance of  $n_i^{(s,u)}$ , into account.

#### D. Distribution of the Magnitude

$\text{var}(n_i^{(s,u)})$  must be determined for  $s = r, h$  to derive the approximate distribution of  $\mathbf{n}^{(s,u)}$ . We denote by  $\stackrel{\mathcal{L}}{=}$  as equality in law [27].

*Lemma 5 (Distribution of Riesz Coefficients):* The DWT coefficients of the original signal and the RTs of Gaussian white noise are distributed as:

$$\mathbf{n}^{(r,u)} \stackrel{\mathcal{L}}{=} \mathbf{Z}^{(r,u)} + O(1/N), \quad \mathbf{Z}^{(r,u)} \sim \mathcal{N}_3\left(\mathbf{0}, \sigma^2 \text{diag}\begin{bmatrix} 1 & a^{(r,u)} & 1 - a^{(r,u)} \end{bmatrix}\right), \quad u = 1, 2, 3, 4. \quad (19)$$

where  $\text{diag}$  denotes a diagonal square matrix,  $a^{(r,u)} = \frac{1}{2}$ ,  $u = 1, 4$ ,  $a^{(r,2)} = \frac{1}{2} + 2 \tan^{-1}\left(\frac{1}{2}\right) + \frac{1}{2} \tan^{-1}(2)$  and  $a^{(r,3)} = \frac{1}{2} - 2 \tan^{-1}\left(\frac{1}{2}\right) - \frac{1}{2} \tan^{-1}(2)$ .

*Proof:* For the proof see Appendix C. ■

*Lemma 6 (Distribution of Riesz Magnitude):* The magnitude square of the DWT of the RTs of Gaussian white noise, denoted  $M_\xi^{(\epsilon,r)2}$ , are distributed as

$$\frac{M_\xi^{(\epsilon,r)2}}{\sigma^2/2} \stackrel{\mathcal{L}}{=} \frac{\mathbf{Z}^{(r,u)T} \mathbf{Z}^{(r,u)}}{\sigma^2} + O(1/N) \stackrel{\mathcal{L}}{=} T_1 + O(1/N), \quad T_1 \sim \chi_1^2 + a^{(r,u)}\chi_1^2 + (1 - a^{(r,u)})\chi_1^2, \quad (20)$$

$j = 1, \dots, J$ ,  $k_1, k_2 = 0, \dots, N_j$ ,  $u = 1, \dots, 4$ , where if  $u = 1, 4$ ,  $T_1$  has distribution

$$f_{T_1}(t) = e^{-t} \frac{2}{\sqrt{\pi}} \int_0^{\sqrt{\frac{t}{2}}} e^{w^2} dw, \quad (21)$$

whilst if  $u = 2, 3$  the moment generating function of  $T_1$  is readily calculable, and we may calculate the probability of obtaining large variates using formulae derived by [28].

*Proof:* For the proof see Appendix C. ■

*Lemma 7 (Distribution of HCT Coefficients):* The DWT coefficients of the HCT of Gaussian white noise are distributed as

$$\mathbf{n}^{(h,u)} \stackrel{\mathcal{L}}{=} \mathbf{Z}^{(h,u)} + O(1/N), \quad \mathbf{Z}^{(h,u)} \sim \mathcal{N}_4\left(\mathbf{0}, \sigma^2 \text{diag}\begin{bmatrix} 1 & 1 & 1 & 1 \end{bmatrix}\right), \quad u = 1, 2, 3, 4. \quad (22)$$

*Proof:* For the proof see Appendix C. ■

Given the approximate joint distribution of the DWT coefficients at  $\xi$  has been determined, it trivially follows that the magnitude is distributed as

$$\frac{M_\xi^{(\epsilon,h)}}{\sigma^2/4} \stackrel{\mathcal{L}}{=} \frac{\mathbf{Z}^{(h,u)T} \mathbf{Z}^{(h,u)}}{\sigma^2} + O(1/N) \stackrel{\mathcal{L}}{=} T_2 + O(1/N), \quad T_2 \sim \chi_1^2 + \chi_1^2 + \chi_1^2 + \chi_1^2 \sim \chi_4^2.$$

### E. Threshold Choice

*Lemma 8 (Riesz Conservative Threshold): Taking*

$$\lambda_K^{(r)2}(l) = 2 \log(K) + 2C \log(\log(K)), \quad (23)$$

it follows that if  $C > -1$ ,

$$P\left(\mathcal{M}_K^{(r)} < \lambda_K^{(r)2}(l)\right) \rightarrow 1. \quad (24)$$

*Proof:* For the proof see Appendix C. ■

From [4] we may note that the RT threshold is thus like that of a  $\chi_1^2$  ( $C > -1$ ) however to ensure that the probability tends to 1 rather than a fixed constant we take  $C = 0$ , rather than  $C = -1$ . Given the normalised marginal magnitudes of the HT components are  $\chi_4^2$  we may use results of [4] to note that

$$\lambda_K^{(h)2}(l) = 2 \log(K) + 2 \log(\log(K)).$$

gives an appropriate threshold. Note that yet again, we expect this to be a conservative threshold, because the wavelet coefficients will be correlated *across*  $\xi$ . As a final step of the procedure we implement cycle-spinning [18, p. 429], which is known to improve mean square error results considerably. Finally for completeness consider implementing hard thresholding in the usual fashion: this will be denoted by  $s = c$ , and we discuss using a single added extra component of  $n_4^{(h,u)}$  when thresholding  $n_1^{(h,u)}$  as a naive extension of ‘analytic’ thresholding, denoted by taking  $s = a$ .

### F. Risk Calculations

To compare the theoretical performance of the threshold estimators proposed in this paper, we calculate the standardized mean square risk at any fixed value of  $\xi$ . We define the standardized risk using any threshold procedure denote by  $s$  for  $\theta_l = \mu_l^{(s,u)}/\sigma$  by

$$R_\theta^{(s)}(\lambda) = \sigma^{-2} E \left[ \left( \hat{\mu}_1^{(s,u)}(\lambda^2) - \mu_1 \right)^2 \right]. \quad (25)$$

If  $s = r$  then we denote by  $r_1$  and  $r_2$  the two different cases that may occur at a given  $\xi$  when  $u = 1, 4$  or  $u = 2, 3$  – the risk will be different in these two cases. This will not happen for  $s = c$  or  $s = h$ . For completeness we here also provide the risk of the ‘analytic’ denoising, as this was not done in Olhede & Walden [15] and corresponds to a special case of the risks determined by Fryzlewicz [14].

*Theorem 1 (The Risk of a Thresholded Coefficient): The standardized risk of an individual coefficient is using threshold strategy  $s = c, a, r, h$  with  $\theta_i = \mu_i^{(s,u)}/\sigma$  and  $R_j(\lambda) = \sum_l (w_l + \theta_l)^2 < \lambda^2$  given by:*

$$\begin{aligned} R_\theta^{(c)}(\lambda) &= 1 + \int_{(w+\theta_1)^2 < \lambda^2} [\theta_1^2 - w^2] \phi(w) dw, \quad R_\theta^{(a)}(\lambda) = 1 + \int_{R_j(\lambda)} [\theta_1^2 - w^2] \phi(w_1) \phi(w_2) d^2 \mathbf{w} \\ R_\theta^{(r)}(\lambda) &= 1 + \frac{1}{\sqrt{a^{(r,u)}(1 - a^{(r,u)})}} \int_{R_j(\lambda)} [\theta_1^2 - w^2] \phi(w_1) \phi(w_2/\sqrt{a^{(r,u)}}) \phi(w_3/\sqrt{1 - a^{(r,u)}}) d^3 \mathbf{w} \\ R_\theta^{(h)}(\lambda) &= 1 + \int_{R_j(\lambda)} [\theta_1^2 - w^2] \phi(w_1) \phi(w_2) \phi(w_3) \phi(w_4) d^4 \mathbf{w}. \end{aligned} \quad (26)$$

*Proof:* The risk of an individual coefficient using standard hard thresholding has been noted by Marron *et al.* [29], whilst the risk of the hyperanalytic thresholds are derived in appendix D, and  $s = a$  is a special case of the bi-variate thresholding investigated by Fryzlewicz [14]. ■

For some examples of signal/noise distributions, the individual risk of a given coefficient is plotted in Figure 2 for the four estimation procedures using the universal threshold. Figure 2 (a) shows the reduced risk of ‘analytic’ thresholding compared to regular thresholding in 1-D when the means of the wavelet coefficient of the signal and of the HT of the signal are equal, and this then provides theoretical justification for the ‘analytic’ denoising procedure. Figures 2 (b), (c) and (d) show the risk associated with thresholding at any index  $\xi$  using either the usual hard thresholding ( $c$ ), ‘analytic’ denoising ( $a$ ), Riesz denoising when  $u = 1, 4$  ( $r_1$ ) and  $u = 2, 3$  ( $r_2$ ) or Hypercomplex denoising ( $h$ ). The risk is calculated with  $K = 256^2$  and using the universal threshold. If the mean of the DWT of the quadrature components is of similar magnitude to the DWT of the signal then the risk is reduced. The greatest weakness of the proposed methods is if the means of the quadrature components are completely disparate from

that of the original signal, as may be noted from figure 2 (c). The results of section III-B indicate that this will not be the case for typical image features. The norm of a signal and its HT are identical, and given the results of section III-B the means are unlikely to be consistently mismatched. Finally if there is no signal present we observe the following result.

*Corollary 1 (The Risk of a Thresholded Coefficient when there is no signal):* The risk of an individual coefficient is using threshold strategy  $s = c, a, r, h$  with  $\theta_{l+1} = 0$  for  $l = 0, \dots, L$  given by:

$$\begin{aligned} R_0^{(c)}(\lambda) &= 1 - \gamma\left(\frac{1}{2}, \frac{1}{2}\lambda^2\right), R_0^{(a)}(\lambda) = e^{-\frac{1}{2}\lambda^2} \left(1 + \frac{1}{2}\lambda^2\right), R_0^{(h)}(\lambda) = e^{-\frac{1}{2}\lambda^2} \left(1 + \frac{1}{2}\lambda^2 + \frac{1}{8}\lambda^4\right) \\ R_0^{(r)}(\lambda) &= 1 + \gamma\left(\frac{1}{2}, \frac{1}{2}\lambda^2\right) - 2\gamma\left(\frac{1}{2}, \lambda^2\right) + 4\lambda^2 \left[\Phi(\sqrt{2}\lambda) - \Phi(\lambda)\right], u = 1, 4. \end{aligned} \quad (27)$$

*Proof:* See appendix D. We denote by  $\gamma(a, x) = \frac{1}{\Gamma(a)} \int_0^x s^{a-1} e^{-s} ds$ . ■

As the representation of the image will be sparse the risk if no signal is present is important. From the corollary, and the asymptotic forms in  $\lambda$  given in appendix D, we may note that the risk at the universal threshold when there is no signal present is of the same order for  $R_0^{(c)}(\sqrt{2\log(K)})$  and  $R_0^{(r)}(\sqrt{2\log(K)})$  even if the coefficient differs in favour of  $R_0^{(c)}(\sqrt{2\log(K)})$  whilst  $R_0^{(h)}(\sqrt{2\log(K\log(K))})$  and  $R_0^{(a)}(\sqrt{2\log(K)})$  correspond to different orders. The thresholds were introduced to improve the estimation of signals that were slightly more spread across coefficients than strictly necessary, but the risk for any coefficient when no signal is present is of similar enough nature to make the difference in estimation negligible (i.e.  $O\left(\frac{\log(K)}{K}\right)$  rather than  $O\left(\frac{1}{K\log(K)}\right)$ ). The examples will substantiate this claim.

#### IV. EXAMPLES

To examine the properties of the proposed methods, we have implemented simulation studies on images that can be retrieved at <http://sipi.usc.edu/database/> (Tiffany and Boat), whilst (Lenna and MRIScan) are downloaded from <http://www-stat.stanford.edu/~wavelab/>. We used LA wavelets length 8. To compare our results, similarly to [5], we also implemented usual hard thresholding and the Wavelet-domain Hidden Markov Model (HMM method) proposed by the Rice group [13], where the software is available at <http://www-dsp.rice.edu/software/>, denoted by  $s = hmm$ . We used the code `hdnoise.m` with default settings, and `daubc9f(8, 'min')`. We implemented the method at several Signal-to-Noise Ratio (SNRs) of 2 (very noisy), 4 and 8 (quite clean), with a set of images, i.e. Lenna ( $512 \times 512$  version), Boat ( $512 \times 512$ ), MRIScan ( $256 \times 256$ ) and the second channel of the colour image of Tiffany ( $512 \times 512$ ). The SNR is (as usual) given by  $SNR^2 = \frac{1}{N^2\sigma^2} \sum \sum q_x^2$ . Table I shows the result over repeated simulations. Reduced Mean Square Errors MSEs, and increased Peak Signal to Noise Ratios PSNRs (using the definition of [30]), are observed when using the proposed method with either hyperanalytic threshold criterion, and the reduction in MSE is of a respectable magnitude compared to variation across replications as the estimated standard deviations in the MSEs are usually considerably smaller. Overall the hypercomplex thresholding procedure is outperforming the Riesz thresholding as well as the other methods, apart from the boat image at high SNRs where the HMM does better. The Hypercomplex method is expected to outperform the Riesz method from the risk calculations, but not perhaps from our discussion in sections II-D and II-E. The Riesz transform may appear more useful as it determines the prevalent direction from the image, whilst the Hypercomplex transform simply decreases the risk in estimation by considering variation associated with the same time-frequency (i.e. 1-D) behaviour in both axes separately, with the second variable treated as fixed. However, whilst the Riesz transform is suitable to use on locally unidirectional structure as discussed in section II-D, the hypercomplex transform treats variation in both axes, and images quite frequently have multidirectional variations present even locally. Some additional analysis of images has also been implemented in [31]. Consider two cuts from reconstructions to further elucidate on these results: see Figure 1 (a)-(h). We show the hypercomplex reconstructions only as the Riesz and hypercomplex reconstructions are similar. Clearly both 1 (c) is more connected than (b), as is (g) to (f), whilst (d) and (h) has much remaining noise to preserve more detail. The proposed method performs quite well. In addition to the SNR's chosen for the full range of images, we implemented the procedure at the SNR chosen by Starck *et al.* [5], namely adding Gaussian noise with a standard deviation of 20 to the raw Lenna image, or using a SNR of 5.58. The PSNR we observed in the noisy image (21.58) is less than theirs (22.13) but that is to be expected in noisy replications. We found for the methods tested in this paper

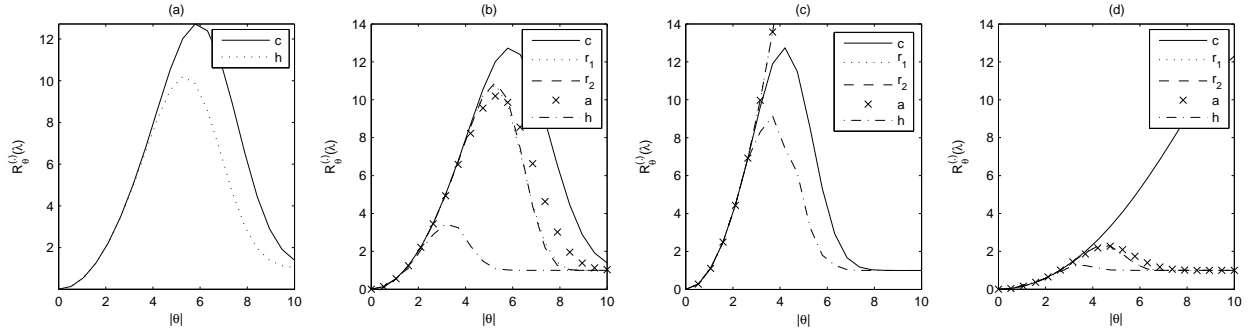


Fig. 2. The risk of hard thresholding compared to ‘analytic’ hard thresholding (a) where the standardised mean of the coefficient is denoted  $\theta = |\theta| \cos(\phi)$ , where  $\phi = \pi/4$ . The risk associated with a thresholded coefficient using standard hard thresholding (solid line), the analytic threshold (crosses), the Riesz thresholds (dotted line and dashed line) or the Hypercomplex threshold (dash-dotted line). In plot (b)  $\theta_1 = \sqrt{\theta_2^2 + \theta_3^2} = |\theta|/\sqrt{2}$  for the Riesz threshold whilst  $\theta_1 = \theta_2 = \theta_3 = \theta_4 = |\theta|/\sqrt{2}$ , for the Hypercomplex threshold. In plot (c)  $\theta_2 = \theta_3 = 0$  for the Riesz threshold whilst  $\theta_1 = \theta_2$  and  $\theta_3 = \theta_4 = 0$  for the Hypercomplex threshold. In plot (d)  $\theta_1 = |\theta| \cos(3\pi/8)$   $\theta_2 = \theta_3 = |\theta| \sin(3\pi/8)/\sqrt{2}$  for the Riesz threshold whilst  $\theta_2 = \theta_1 = |\theta| \cos(3\pi/8)$  and  $\theta_3 = \theta_4 = |\theta| \sin(3\pi/8)$  for the Hypercomplex threshold. For the ‘analytic’ procedure we use  $\theta_4$  as the second component.

that averaged over 100 replications  $s = c$  (29.22),  $s = r$  (30.12),  $s = h$  (30.93) and  $s = hmm$  (30.48), where they obtained for  $s = c$  (28.35) and  $s = hmm$  (30.80). Starck obtained PSNRs between 29.99 to 31.95 by using local ridgelets and curvelets. Clearly the hypercomplex denoising performs on par with the algorithms suggested, and in addition the proposed procedure is both cheap to implement and extremely simple to code.

TABLE I

THE AVERAGE RESULTS OVER 50 RUNS. THE SYMMLET WAVELETS (OR LA WAVELETS WERE USED) AND  $J = 3$ .

Example (SNR)	Boat (2/4/8)	Lena (2/4/8)	Tiffany (2/4/8)	MRIScan (2/4/8)
average MSE (n) $\times 10^{-8}$	95.4/23.8/6.0	95.4/23.8/6.0	95.4/23.8/6.0	381.5/95.4/23.8
sd MSE (n) $\times 10^{-10}$	26.3/6.6/1.6	26.3/6.6/1.6	26.3/6.6/1.6	187.9/47.0/11.7
PSNR (n)	11.36/17.38/23.40	10.67/16.69/22.71	7.65/13.67/19.69	16.96/22.98/29.00
average MSE (c) $\times 10^{-8}$	8.0/4.4/2.2	5.16/2.6/1.1	3.8/2.2/1.0	56.9/24.5/9.1
sd MSE (c) $\times 10^{-10}$	6.7/3.2/1.2	7.3/2.3/0.8	4.9/2.5/0.8	110.0/43.4/15.3
PSNR (c)	22.12/24.68/27.70	23.34/26.34/29.88	21.64/24.02/27.39	25.22/28.88/33.18
average MSE (r) $\times 10^{-8}$	7.1/3.8/1.8	4.5/2.2/0.93	3.4/1.9/0.8	46.9/19.7/7.3
sd MSE (r) $\times 10^{-10}$	6.2/2.7/1.0	5.6/2.1/0.7	3.8/2.4/0.7	89.2/36.9/11.6
PSNR (r)	22.63/25.39/28.49	23.96/27.09/30.77	22.09/24.73/28.00	26.06/29.82/34.17
average MSE (h) $\times 10^{-8}$	<b>6.3/3.2/1.5</b>	<b>3.9/1.9/0.77</b>	<b>3.1/1.7/0.8</b>	<b>40.4/16.6/6.1</b>
sd MSE (h) $\times 10^{-10}$	6.6/2.2/0.9	4.6/2.0/0.6	3.3/2.3/0.6	79.1/31.0/8.7
PSNR (h)	<b>23.14/26.04/29.2</b>	<b>24.56/27.74/31.58</b>	<b>22.48/25.06/28.55</b>	<b>26.71/30.57/34.92</b>
average MSE (hmm) $\times 10^{-8}$	6.7/ <b>3.0/1.4</b>	6.3/2.1/0.85	5.4/1.8/ <b>0.8</b>	48.9/20.9/7.5
sd MSE (hmm) $\times 10^{-10}$	11.3/1.8/0.7	16.0/2.3/0.5	16.4/1.2/2.7	82.1/56.9/7.4
PSNR (hmm)	22.90/ <b>26.38/29.62</b>	22.48/27.31/31.19	20.12/24.78/28.26	25.88/29.57/34.01

## V. CONCLUSIONS

This paper has proposed a new thresholding strategy for estimating decomposition coefficients, and has in particular implemented the strategy with the discrete separable DWT. We have determined the stochastic properties of the decomposition of the noise, and both the deterministic (for some stylized image features) and stochastic properties of the suggested new thresholding criterion. We established universal thresholds. We calculated the risk theoretically, and for some specific choices of the mean provided plots of the risk showing that the proposed methods outperform standard denoising theoretically. We implemented the procedure on several examples at several SNRs, comparing the methods with the Hidden-Markov-Model used by the Rice group as well as standard hard thresholding, and found that the proposed algorithms offered improvements in most cases. Given the simplicity in implementation, and visually pleasing reconstructions, hyperanalytic denoising methods offer a computationally cheap improvement to existing methodology, as well as offers insight into 2-D variational structure.

## ACKNOWLEDGMENTS

SO would like to express her thanks to the anonymous referees for the many helpful suggestions that substantially improved the paper, as well as her understanding of the topic. SO would also like to thank Professor Andrew Walden for introducing her to this research area, and gratefully acknowledges financial support from EPSRC (UK). SO gratefully acknowledges usage of WaveLab routines and the data sets.

## REFERENCES

- [1] D. L. Donoho and I. M. Johnstone, "Ideal spatial adaption via wavelet shrinkage," *Biometrika*, vol. 81, pp. 425–455, 1994.
- [2] D. L. Donoho and I. M. Johnstone, "Adapting to unknown smoothness via wavelet shrinkage," *Journal of the Am. Stat. Assoc.*, vol. 90, pp. 1200–1224, 1995.
- [3] C. Stein, "Estimation of the Mean of a Multivariate Normal Distribution," *The Annals of Statistics*, vol. 9, pp. 1135–1151, 1981.
- [4] T. R. Downie and B. W. Silverman, "The discrete multiple wavelet transform and threshold methods," *IEEE Trans. Signal Processing*, vol. 46, pp. 2558–2561, 1998.
- [5] J. L. Starck, E. J. Candès and D. L. Donoho, "The curvelet transform for image denoising," *IEEE Trans. Image Processing*, vol. 11, pp. 670–684, 2002.
- [6] D. L. Donoho, "Wedgelets: Nearly Minimax Estimation of Edges," *Annals of Statistics*, vol. 27, pp. 859–897, 1999.
- [7] E. le Pennec and S. Mallat, "Sparse Geometric Image Representations," *IEEE Trans. Image Processing*, vol. 14, pp. 423–438, 2005.
- [8] M. Jansen and A. Bultheel, "Empirical Bayes Approach to Improve Wavelet Thresholding for Image Noise Reduction," *J. Am. Stat. Assoc.*, vol. 96, pp. 629–639, 2001.
- [9] "Needles and Hay in Haystacks: Empirical Bayes Estimates of Possibly Sparse Sequences," I. M. Johnstone and B. W. Silverman, *Annals of Statistics*, vol. 32, pp. 1594–1649, 2004.
- [10] T. Cai and B. W. Silverman, "Incorporating information on neighbouring coefficients into wavelet estimation", *Sankhyā Ser. B*, vol. 63, pp. 127–148, 2001.
- [11] P. L. Dragotti and M. Vetterli, "Wavelet footprints: theory, algorithms and applications", *IEEE Trans. Signal Processing*, vol. 51, pp. 1306–1323, 2003.
- [12] A. Pižurica and W. I. Philips and I. Lemahieu and M. Acheroy", "A joint inter- and intrascale statistical model for Bayesian wavelet based image denoising," *IEEE Trans on Image Proc.*, vol. 11, pp. 545–557, 2002.
- [13] M. S. Crouse, R. Nowak and R. G. Baraniuk, "Wavelet-Based Statistical Signal Processing Using Hidden Markov Models," *IEEE Trans. Signal Processing*, vol. 46, pp. 886–902, 1998.
- [14] P. Fryzlewicz, P.", "Bivariate Hard Thresholding in Wavelet Function Estimation," *Statistica Sinica*, to appear.
- [15] S. C. Olhede and A. T. Walden, "'Analytic' wavelet thresholding," *Biometrika*, vol. 91, pp. 955–973, 2004.
- [16] S. C. Olhede and G. Metikas, The Hyperanalytic Wavelet Transform, Imperial College Statistics Section, TR-06-02, arXiv math.ST/0605623, 2006.
- [17] S. Mallat. *A wavelet tour of signal processing*, 2nd ed, Academic Press, New York, USA. 1999.
- [18] D. B. Percival and A. T. Walden, *Wavelet Methods for Time Series Analysis*, Cambridge University Press, Cambridge, UK, 2000.
- [19] B. Boashash, Note on the use of the Wigner distribution for time-frequency signal analysis, *IEEE Trans. Accoust., Speech, Signal Proc.*, vol. 36, pp. 1518–1521, 1988.
- [20] G. Metikas and S. C. Olhede, Multiple Monogenic Morse Wavelets, Imperial College Statistics Section, TR-05-02, revised version to appear in *IEEE Trans. Signal Proc.*, arXiv math.ST/0511324, 2005.
- [21] S. L. Hahn, *Hilbert transforms in signal processing*, Boston, MA, US, Artech House, 1996.
- [22] L. A. Vese & S. J. Osher, "Modeling Textures with Total Variation Minimization and Oscillating Patterns in Image Processing," *J. Scien. Comp.*, vol. 19, pp. 553–572, 2003.
- [23] M. Nielsen, "On the construction and frequency localisation of finite orthogonal quadrature filters," *J. of Approximation Theory*, vol. 108, pp. 36–52, 2001.
- [24] R. A. Gopinath, "Phaselets of Framelets," *IEEE Trans. Signal Processing*, vol. 53, pp. 1794–1806, 2005.
- [25] I. M. Johnstone and B. W. Silverman, "Wavelet Threshold Estimators for Data with Correlated Noise," *J. Royal Statistical Society B*, vol. 59, pp. 319–351, 1997.
- [26] R. L. Dykstra, "Product inequalities involving the multivariate normal distribution," *J. Am. Stat. Assoc.*, vol. 75, pp. 646–650, 1980.
- [27] T. S. Ferguson, *A Course in Large Sample Theory*, Chapman & Hall/CRC, London, UK, 1996.
- [28] A. Grad and A. Solomon, "Distribution of Quadratic forms and some applications," *The Annals of Mathematical Statistics*, vol. 26, pp. 464–477, 1955.
- [29] J. S. Marron and S. Adak, "Exact Risk Analysis Of Wavelet Regression," *J. Computational and Graphical Statistics*, vol. 7, pp. 278–309, 1998.
- [30] S. Barber and G. P. Nason, "Denoising Real Images Using Complex-Valued Wavelets," in *Stochastic Geometry, Biological Structure and Images*, Department of Statistics, University of Leeds, pp. 91–95, 2003.
- [31] S. C. Olhede, "Hyperanalytic denoising," to appear at *IEEE International Conference on Image Processing (ICIP)*, Atlanta, Georgia, October, 2006.
- [32] I. I. Gikhman and A. V. Skorokhod, *The Theory of Stochastic Processes*, Springer Classics in Mathematics, London, UK, 1980.

## APPENDIX

## A: DIGITAL IMPLEMENTATION

For future reference the Discrete Fourier Transform (DFT) and its inverse are given with  $\Omega = [-\frac{1}{2\Delta x}, \frac{1}{2\Delta x}]^2$  by:

$$Q_D(f_1, f_2) = \Delta x^2 \sum_{x_1=0}^{N-1} \sum_{x_2=0}^{N-1} q(x_1\Delta x, x_2\Delta x) e^{-2i\pi \mathbf{f}^T \mathbf{x} \Delta x}, \quad q_{\mathbf{x}} = \int \int_{\Omega} Q_D(\mathbf{f}) e^{2i\pi \mathbf{f}^T \mathbf{x}} d^2 \mathbf{f}. \quad (28)$$

Note that the value of  $Q_D(f_1, f_2)$  at frequencies  $f_1 \in \left\{ \frac{N/2}{N\Delta x}, \frac{N/2+1}{N\Delta x}, \frac{N/2+2}{N\Delta x}, \dots, \frac{N-1}{N\Delta x} \right\}$  is equivalent to the value of  $Q_D(f_1, f_2)$  at frequencies  $f_1 \in \left\{ -\frac{N/2}{N\Delta x}, -\frac{N/2-1}{N\Delta x}, -\frac{N/2-2}{N\Delta x}, \dots, -\frac{1}{N\Delta x} \right\}$ , and the equivalent statement *mutatis mutandis* hold for  $f_2$ . Let  $Q_{D,\mathbf{u}} = Q_D(\frac{u_1}{N\Delta x}, \frac{u_2}{N\Delta x})$ .

$$\check{q}_{\mathbf{x}}^{(s,l)} = \int \int_{\Omega} Q_D(\mathbf{f}) V_D^{(s,l)}(\mathbf{f}) e^{2i\pi \mathbf{f}^T \mathbf{x} \Delta x} d^2 \mathbf{f} = \frac{1}{\Delta x^2 N^2} \sum_{u_1=0}^{N-1} \sum_{u_2=0}^{N-1} Q_{D,\mathbf{u}} V_{D,\mathbf{u}}^{(s,l)} e^{2i\pi \mathbf{u}^T \mathbf{x}} + O\left(\frac{1}{N}\right),$$

$s = r, h$ , and  $l = 1, \dots, L$ , where we define:

$$V_{D,\mathbf{u}}^{(r,1)} = \begin{cases} 0 & \text{if } u_1 = 0 \\ \frac{u_1/(N\Delta x)}{i\sqrt{(u_1/(N\Delta x))^2 + (l_2/(N\Delta x))^2}} & \text{if } u_1 = 1, \dots, N/2, u_2 = 1, \dots, N/2 \\ \frac{(N-u_1)/(N\Delta x)}{i\sqrt{((N-u_1)/(N\Delta x))^2 + (u_2/(N\Delta x))^2}} & \text{if } u_1 = N/2 + 1, \dots, N-1, u_2 = 1, \dots, N/2 \\ \frac{u_1/(N\Delta x)}{i\sqrt{(u_1/(N\Delta x))^2 + ((N-u_2)/(N\Delta x))^2}} & \text{if } u_1 = 1, \dots, N/2, u_2 = N/2 + 1, \dots, N-1 \\ \frac{(N-u_1)/(N\Delta x)}{i\sqrt{((N-u_1)/(N\Delta x))^2 + ((N-u_2)/(N\Delta x))^2}} & \text{if } u_1 = N/2 + 1, \dots, N-1, u_2 = 1, \dots, N/2 \end{cases}.$$

We define  $V_{D,\mathbf{u}}^{(r,2)}$  in an analogous fashion. The digital definition of the filters corresponding to the three Hypercomplex components are defined via:

$$V_{D,\mathbf{u}}^{(h,l)} = \begin{cases} 0 & \text{if } u_l = 0, N/2 \\ -i & \text{if } u_l = 1, \dots, N/2 - 1 \\ i & \text{if } u_l = N/2 + 1, \dots, N-1 \end{cases}, \quad l = 1, 2,$$

and  $V_{D,\mathbf{u}}^{(h,3)} = V_{D,\mathbf{u}}^{(h,1)} V_{D,\mathbf{u}}^{(h,2)}$ . Implementing the discrete HCT introduces an error term of  $O(1/N)$ .

## B: DWT OF TYPICAL DETERMINISTIC IMAGE FEATURES

We use period boundary treatment (see Mallat [17][p. 282–292]) when implementing the MODWT, so that with  $\xi_{\mathbf{x}} = [j, u, \mathbf{x}]^T$ :

$$\tilde{W}_{\xi_{\mathbf{x}}}^{(q)} = \int \int_{\Omega} \tilde{H}_{j,u}(\mathbf{f}) Q_D(\mathbf{f}) e^{2i\pi \mathbf{f}^T \mathbf{x}} d^2 \mathbf{f}. \quad (29)$$

*Proof of Lemma 1*

By direct calculation using equation (29):

$$\begin{aligned} \mu_1 &= 2^j \int \int_{\Omega} \frac{a_t(\mathbf{x}_0)}{2} (\delta(\mathbf{f}_0 - \mathbf{f}) + \delta(\mathbf{f}_0 + \mathbf{f})) \tilde{H}_{j,u}(\mathbf{f}) e^{2i\pi \mathbf{f}^T \mathbf{x}} d^2 \mathbf{f} + O\left(\frac{1}{N}\right) \\ &= 2^j a_t(\mathbf{x}_0) \left| \tilde{H}_{j,u}(\mathbf{f}_0) \right| \cos(2\pi(\mathbf{f}_0^T \mathbf{x} - \varphi_{j,u}(\mathbf{f}_0))) + O\left(\frac{1}{N}\right) \\ \mu_2^{(r,u)} &= 2^j \int \int_{\Omega} \frac{a(\mathbf{x}_0)}{2} (-i) \text{sgn}(f_1) \frac{f_1}{f} (\delta(\mathbf{f}_0 - \mathbf{f}) + \delta(\mathbf{f}_0 + \mathbf{f})) \tilde{H}_{j,u}(\mathbf{f}) e^{2i\pi \mathbf{f}^T \mathbf{x}} d^2 \mathbf{f} \\ &\quad + O\left(\frac{1}{N}\right) = 2^j a_t(\mathbf{x}_0) \left| \tilde{H}_{j,u}(\mathbf{f}_0) \right| \cos(\phi_0) \sin(2\pi(\mathbf{f}_0^T \mathbf{x} - \varphi_{j,u}(\mathbf{f}_0))) + O\left(\frac{1}{N}\right), \end{aligned}$$

and similarly  $\mu_3^{(r,u)} = 2^j a_t(\mathbf{x}_0) \left| \tilde{H}_{j,u}(\mathbf{f}_0) \right| \sin(\phi_0) \sin(2\pi(\mathbf{f}_0^T \mathbf{x} - \varphi_{j,u}(\mathbf{f}_0))) + O\left(\frac{1}{N}\right)$ , whilst  $\rho_1$  is introduced when  $\left| \tilde{H}_{j,u}(\mathbf{f}_0) \right|$  is approximated by an exact band-pass structure.

*Proof of Lemma 2*

By direct calculation, using equation (29), it follows that:

$$\begin{aligned}\mu_2^{(h,u)} &= 2^j \int \int_{\Omega} \frac{a(\mathbf{x}_0)}{2} (-i) \text{sgn}(f_1) (\delta(\mathbf{f}_0 - \mathbf{f}) + \delta(\mathbf{f}_0 + \mathbf{f})) \tilde{H}_{j,u}(\mathbf{f}) e^{2i\pi \mathbf{f}^T \mathbf{x}} d^2 \mathbf{f} \\ &\quad + O\left(\frac{1}{N}\right) = 2^j a_t(\mathbf{x}_0) \left| \tilde{H}_{j,u}(\mathbf{f}_0) \right| \sin(2\pi(\mathbf{f}_0^T \mathbf{x} - \varphi_{j,u}(\mathbf{f}_0))) + O\left(\frac{1}{N}\right).\end{aligned}$$

Similarly  $\mu_3^{(h,u)} = 2^j a_t(\mathbf{x}_0) \left| \tilde{H}_{j,u}(\mathbf{f}_0) \right| \sin(2\pi(\mathbf{f}_0^T \mathbf{x} - \varphi_{j,u}(\mathbf{f}_0))) + O\left(\frac{1}{N}\right)$  and

$$\mu_4^{(h,u)} = 2^j a_t(\mathbf{x}_0) \left| \tilde{H}_{j,u}(\mathbf{f}_0) \right| \cos(2\pi(\mathbf{f}_0^T \mathbf{x} - \varphi_{j,u}(\mathbf{f}_0))) + O\left(\frac{1}{N}\right).$$

*Proof of Lemma 3*

Assume for simplicity of exposition,  $0 < \theta \leq \pi/2$ , but with the necessary notational changes no such restriction needs to be made. Define the rotation matrix by  $\mathbf{r}_\theta = [\cos(\theta) \ -\sin(\theta), \sin(\theta) \ \cos(\theta)]$ , and the change of variable given by:  $\boldsymbol{\nu} = \mathbf{f}(-\theta) = \mathbf{r}_{-\theta} \mathbf{f}$ . We assume that  $A_e(f)$  decays for large frequencies, and consider  $A_e(f) = A_e(0)\delta(f - 0)$ . For example if  $a_e(x) = I(x \in [0, L]) / \sqrt{L}$  then  $A_e(f) = \frac{e^{-i\pi f L} \sin(\pi f L)}{\pi f \sqrt{L}}$ , that as  $L \rightarrow \infty$   $A_e(f)$  will concentrate to  $f = 0$ .  $E(\mathbf{f}) = \frac{A_e(-f_2(\theta)/\sin(\theta))}{|\sin(\theta)|} e^{-2\pi i c f_2/\sin(\theta)}$ , and with  $\Omega(\theta)$  the rotated by  $\theta$  version of  $\Omega$ , with  $\nu_{1,\min} = \min_{\boldsymbol{\nu}} \nu = \sec(\theta) f_1 I(|\tan(\theta) f_1| < \frac{1}{2\Delta x}) I(|f_1| < \frac{1}{2\Delta x})$  and  $\nu_{1,\max} = \max_{\boldsymbol{\nu}} \nu = \sec(\theta) f_1 I(|\tan(\theta) f_1| < \frac{1}{2\Delta x}) I(|f_1| < \frac{1}{2\Delta x})$ , then

$$\begin{aligned}\mu_1 &= 2^j \int \int_{\Omega} \tilde{H}_{j,u}(\mathbf{f}) \frac{A_e(-f_2(\theta)/\sin(\theta))}{|\sin(\theta)|} e^{-2\pi i c f_2/\sin(\theta)} e^{2\pi i \mathbf{f}^T \mathbf{x}} d^2 \mathbf{f} + O\left(\frac{1}{N}\right) \\ &= \int \int_{\Omega(\theta)} \tilde{H}_{j,u}(\mathbf{r}_{-\theta} \boldsymbol{\nu}) \frac{2^j A_e(-\nu_2/\sin(\theta))}{|\sin(\theta)|} e^{-2\pi i ((\sin(\theta)\nu_1 + \cos(\theta)\nu_2)c/\sin(\theta) - \boldsymbol{\nu}^T \mathbf{r}_{-\theta} \mathbf{x})} d^2 \boldsymbol{\nu} + O\left(\frac{1}{N}\right) \\ &\stackrel{(1)}{=} 2^j A_e(0) \int_{\nu_{1,\min}}^{\nu_{1,\max}} \tilde{H}_{j,u}(\cos(\theta)\nu_1, \sin(\theta)\nu_1) e^{2\pi i \nu_1 (\cos(\theta)x_1 + \sin(\theta)x_2 - c)} d\nu_1 + \rho_3 + O\left(\frac{1}{N}\right).\end{aligned}$$

The approximation in (1) relies on  $A_e(\cdot)$  taking the form of a  $\delta$  distribution contribution, i.e.  $a_e(\cdot)$  constant over a large spatial domain, but a slowly varying  $a_e(\cdot)$  will approximately yield the same result. Also we may find approximate descriptions for decomposition of the RTs, namely with  $f = \sqrt{f_1^2 + f_2^2}$  and  $\nu = \sqrt{\nu_1^2 + \nu_2^2}$ :

$$\begin{aligned}\mu_2^{(r,u)} &= 2^j \int \int_{\Omega} \tilde{H}_{j,u}(\mathbf{f}) (-i) f_1 / f \frac{A_e(-f_2(\theta)/\sin(\theta))}{|\sin(\theta)|} e^{-2\pi i c f_2/\sin(\theta)} e^{2\pi i \mathbf{f}^T \mathbf{x}} d^2 \mathbf{f} \\ &\quad + O\left(\frac{1}{N}\right) \stackrel{(1)}{=} A_e(0) 2^j (-i) \cos(\theta) \int_{\nu_{1,\min}}^{\nu_{1,\max}} \tilde{H}_{j,u}(\cos(\theta)\nu_1, \sin(\theta)\nu_1) e^{2\pi i \nu_1 (\cos(\theta)x_1 + \sin(\theta)x_2 - c)} \\ &\quad \times \text{sgn}(\nu_1) d\nu_1 + \rho'_4 + O\left(\frac{1}{N}\right) = \cos(\theta) U_{\xi}^{(e)} + \rho'_4 + O\left(\frac{1}{N}\right).\end{aligned}\tag{30}$$

The approximation in (1) relies on the envelope being constant in the spatial domain – a slowly varying envelope will thus only approximately yield the same value, and this introduces an error term  $\rho'_4$ . Similarly it transpires that  $\mu_3^{(r,u)} = \sin(\theta) U_{\xi}^{(e)} + \rho_7 + O\left(\frac{1}{N}\right)$ , and thus the result follows, with a new error term  $\rho_5$ .

*Proof of Lemma 4*

Also we may find approximate descriptions for the Hypercomplex components, namely:

$$\begin{aligned}\mu_2^{(h,u)} &= 2^j \int \int_{\Omega} \tilde{H}_{j,u}(\mathbf{f}) (-i) \text{sgn}(f_1) \frac{A_e(-f_2(\theta)/\sin(\theta))}{|\sin(\theta)|} e^{-2\pi i c f_2/\sin(\theta)} e^{2\pi i \mathbf{f}^T \mathbf{x}} d^2 \mathbf{f} \\ &\quad + O\left(\frac{1}{N}\right) \stackrel{(1)}{=} 2^j (\pm 1) (-i) \int_{\nu_{1,\min}}^{\nu_{1,\max}} \tilde{H}_{j,u}(\cos(\theta)\nu_1, \sin(\theta)\nu_1) e^{-2\pi i c \nu_1} e^{2\pi i \nu_1 (\cos(\theta)x_1 + \sin(\theta)x_2)} \\ &\quad \times \text{sgn}(\nu_1) d\nu_1 + \rho_8 + O\left(\frac{1}{N}\right) = (\pm) U_{\xi}^{(e)} + \rho_8 + O\left(\frac{1}{N}\right),\end{aligned}\tag{31}$$

where the value of  $\pm$  depends on the value of  $\theta$ . Similarly it transpires that  $\mu_3^{(h,u)} = \pm U_{\xi}^{(e)} + \rho_9 + O\left(\frac{1}{N}\right)$  and  $\mu_4^{(h,u)} = \pm W_{\xi}^{(e)} + \rho_{10} + O\left(\frac{1}{N}\right)$ .  $\rho_8$ ,  $\rho_9$  and  $\rho_{10}$  are constants depending on the variability of  $a_e(\mathbf{x})$ .



## C: STATISTICS OF THE NORMAL VECTOR

For simplicity set  $\Delta x = 1$  when deriving the statistical properties of the coefficients.  $\epsilon_{\mathbf{x}}$  has a spectral representation:  $\epsilon_{\mathbf{x}} = \int_{\Omega} dZ_{\epsilon}(\mathbf{f}) e^{2\pi i \mathbf{f}^T \mathbf{x}}$ , where  $Z_{\epsilon}(\mathbf{f})$  is a complex-valued orthogonal increment process, see [32][p. 244], i.e.  $E(dZ_{\epsilon}(\mathbf{f}) dZ_{\epsilon}^*(\mathbf{f}')) = 0$  if  $\mathbf{f} \neq \mathbf{f}'$ . The DWT is represented by subsampling the MODWT:

$\tilde{\mathbf{n}}^{(s,u)} = [\tilde{W}_{\xi_{\mathbf{x}}}^{(\epsilon,s,0)}, \dots, \tilde{W}_{\xi_{\mathbf{x}}}^{(\epsilon,s,L)}]^T$ , noting that:

$$\text{cov} \left( n^{(s,l_1,u)}, n^{(s,l_2,u)} \right) = 2^{2j} \text{cov} \left( \tilde{n}^{(s,l_1,u)}, \tilde{n}^{(s,l_2,u)} \right). \quad (32)$$

We have that  $\tilde{n}_1^{(s,u)} = \int_{\Omega} \tilde{H}_{j,u}(\mathbf{f}) dZ_{\epsilon}(\mathbf{f}) e^{2\pi i \mathbf{f}^T \mathbf{x}}$ ,  $\tilde{H}_{j,1}(\mathbf{f}) = \tilde{H}_j(f_1) \tilde{H}_j(f_2)$ ,  $\tilde{H}_{j,2}(\mathbf{f}) = \tilde{H}_j(f_1) \tilde{G}_j(f_2)$ ,  $\tilde{H}_{j,3}(\mathbf{f}) = \tilde{G}_j(f_1) \tilde{H}_j(f_2)$  and  $\tilde{H}_{j,4}(\mathbf{f}) = \tilde{G}_j(f_1) \tilde{G}_j(f_2)$ . We approximate the magnitude of the wavelet filters as exact bandpass filters - see for example Nielsen [23] for a discussion of such approximations, and optimal filters to use. That is:

$$\left| \tilde{H}_j(f) \right|^2 = \begin{cases} 1 & \text{if } |f| \in \left[ \frac{1}{2^{j+1}}, \frac{1}{2^j} \right) \\ 0 & \text{if } |f| \in \left[ \frac{1}{2^{j+1}}, \frac{1}{2^j} \right) \end{cases}, \quad \left| \tilde{G}_j(f) \right|^2 = \begin{cases} 1 & \text{if } f \in \left( -\frac{1}{2^{j+1}}, \frac{1}{2^{j+1}} \right) \\ 0 & \text{if } |f| \in \left[ \frac{1}{2^{j+1}}, \frac{1}{2^j} \right) \end{cases}. \quad (33)$$

$$\text{Var} \left( n_1^{(s,u)} \right) = 2^{2j} E \left( \int \int_{\Omega} \int \int_{\Omega} \tilde{H}_{j,u}(\mathbf{f}) dZ_{\epsilon}(\mathbf{f}) e^{2\pi i (\mathbf{f} - \mathbf{f}')^T \mathbf{x}} \tilde{H}_{j,u}^*(\mathbf{f}') dZ_{\epsilon}^*(\mathbf{f}') \right) = \sigma^2$$

*Proof of Propositions 1 & 2*

$$\begin{aligned} \text{var} \left( n_l^{(s,u)} \right) &= 2^{2j} E \left( \int \int_{\Omega} \int \int_{\Omega} \tilde{H}_{j,u}(\mathbf{f}) \tilde{H}_{j,u}^*(\mathbf{f}') e^{2\pi i (\mathbf{f}^T - \mathbf{f}'^T) \mathbf{x}} dZ_{\epsilon}(\mathbf{f}) dZ_{\epsilon}^*(\mathbf{f}') \right) \\ &= 2^{2j} \int_{\Omega} \left| V_D^{(s,l)}(\mathbf{f}) \right|^2 \left| \tilde{H}_{j,u}(\mathbf{f}) \right|^2 d^2 \mathbf{f} \\ &= 2^{2j} \int_{\Omega} \left| V^{(s,l)}(\mathbf{f}) \right|^2 \left| \tilde{H}_{j,u}(\mathbf{f}) \right|^2 d^2 \mathbf{f} + O \left( \frac{1}{N} \right) \equiv a_l^{(s,u)} + O \left( \frac{1}{N} \right), \end{aligned}$$

The latter defining  $a_l^{(s,u)}$ , to be explicitly determined for  $s = r, h$  and  $u = 1, 2, 3, 4$ . For  $l_1 \neq l_2$ , where  $l_1, l_2 \neq 0$ , we determine that:

$$\begin{aligned} \text{cov} \left( n_{l_1}^{(s,u)}, n_{l_2}^{(s,u)} \right) &= 2^{2j} \int_{\Omega} \sigma^2 V_D^{(s,l_1)}(\mathbf{f}) V_D^{(s,l_2)*}(\mathbf{f}) \left| \tilde{H}_{j,u}(\mathbf{f}) \right|^2 d^2 \mathbf{f} \\ &= 2^{2j} \sum_{\mathbf{x}} \sum_{\xi_{\mathbf{x}}} \check{h}_{\xi_{\mathbf{x}}}^{(s,l_1)} \check{h}_{\xi_{\mathbf{x}}}^{(s,l_2)*} + O \left( \frac{1}{N} \right) = O \left( \frac{1}{N} \right), \end{aligned} \quad (34)$$

by property 1 as  $\tilde{h}_{\xi_{\mathbf{x}}}$  is separable. Therefore

$$\begin{aligned} E \left( \sum_{l=1}^L n_l^{(s,u)2} \right) &= 2^{2j} \int_{\Omega} \sum_{l=1}^L \left| V^{(s,l)}(\mathbf{f}) \right|^2 \sigma^2 \left| \tilde{H}_{j,u}(\mathbf{f}) \right|^2 d^2 \mathbf{f} + O \left( \frac{1}{N} \right) \\ &= 2^{2j} \int_{\Omega} C_L^{(s)} \sigma^2 \left| \tilde{H}_{j,u}(\mathbf{f}) \right|^2 d^2 \mathbf{f} + O \left( \frac{1}{N} \right) \\ &= C_L^{(s)} \text{var} \left( n_1^{(s,u)2} \right) + O \left( \frac{1}{N} \right). \end{aligned} \quad (35)$$

Hence the total energy of the noise associated with the total magnitude square of the added quadrature components is a constant times the variance of the original signal.

$$\begin{aligned} \text{cov} \left( n_l^{(s,u)} n_1^{(s,u)} \right) &= 2^{2j} E \left( \int \int_{\Omega} \int \int_{\Omega} \tilde{H}_{j,u}(\mathbf{f}) \tilde{H}_{j,u}^*(\mathbf{f}') dZ_{\epsilon}(\mathbf{f}) \right. \\ &\quad \left. dZ_{\epsilon}^*(\mathbf{f}') e^{2\pi i (\mathbf{f} - \mathbf{f}')^T \mathbf{x}} \right) = \int_{\Omega} \sigma^2 V^{(s,l)}(\mathbf{f}) \tilde{H}_{j,u}(\mathbf{f}) \tilde{H}_{j,u}^*(\mathbf{f}) d^2 \mathbf{f} \\ &\quad + O \left( \frac{1}{N} \right) = \sum_{\mathbf{x}} \check{h}_{\xi_{\mathbf{x}}}^{(s,l)} \check{h}_{\xi_{\mathbf{x}}}^{(s,l)} + O \left( \frac{1}{N} \right) = O \left( \frac{1}{N} \right). \end{aligned}$$

*Proof of Lemma 5*

Given the noise was Gaussian and zero-mean we only need to determine the second order structure to deduce the Lemma, using equation (32).

$$\begin{aligned}
\text{Var} \left( n_2^{(r,1)} \right) &= 2^{2j} E \left( \int \int_{\Omega} \int \int_{\Omega} \tilde{H}_j(f_1) \tilde{H}_j(f_2) dZ_{\epsilon}(\mathbf{f}) e^{2\pi i(\mathbf{f}-\mathbf{f}')^T \mathbf{x}} \right. \\
&\quad \left. \frac{f_1}{\sqrt{f_1^2 + f_2^2}} \tilde{H}_j^*(f'_1) \tilde{H}_j^*(f'_2) dZ_{\epsilon}(\mathbf{f}') \frac{f'_1}{\sqrt{f_1'^2 + f_2'^2}} \right) + O \left( \frac{1}{N} \right) \\
&= 2^{2j} \sigma^2 \int \int_{\Omega} \frac{f_1^2}{f_1^2 + f_2^2} \left| \tilde{H}_j(f_1) \right|^2 \left| \tilde{H}_j(f_2) \right|^2 d^2 \mathbf{f} + O \left( \frac{1}{N} \right). \\
\text{Var} \left( n_3^{(r,1)} \right) &= 2^{2j} \sigma^2 \int \int_{\Omega} \frac{f_2^2}{f_1^2 + f_2^2} \left| \tilde{H}_j(f_1) \right|^2 \left| \tilde{H}_j(f_2) \right|^2 d^2 \mathbf{f} + O \left( \frac{1}{N} \right).
\end{aligned}$$

Hence it follows that

$$\text{Var} \left( n_2^{(r,1)} \right) + \text{Var} \left( n_3^{(r,1)} \right) = \text{Var} \left( n_1^{(r,1)} \right), \quad (36)$$

and the two variances are obviously equal. Also:

$$\text{Var} \left( n_2^{(r,u)} \right) + \text{Var} \left( n_3^{(r,u)} \right) = \text{Var} \left( n_1^{(r,u)} \right), \quad u = 1, 2, 3, 4, \quad (37)$$

$$\begin{aligned}
\text{Var} \left( n_2^{(r,1)} \right) &= 2^{2j+2} \sigma^2 \int_{\frac{1}{2^{j+1}}}^{\frac{1}{2^j}} \int_{\frac{1}{2^{j+1}}}^{\frac{1}{2^j}} \frac{f_1^2}{f_1^2 + f_2^2} df_2 df_1 + O \left( \frac{1}{N} \right) \\
&= 2^{2j+2} \sigma^2 \int_{\frac{1}{2^{j+1}}}^{\frac{1}{2^j}} f_1 \left[ \tan^{-1} \left( \frac{f_2}{f_1} \right) \right]_{f_2=\frac{1}{2^{j+1}}}^{\frac{1}{2^j}} df_1 + O \left( \frac{1}{N} \right) \\
&= \sigma^2 \left( \frac{1}{4} - \frac{1}{8} \tan^{-1}(2) + \frac{1}{2} \tan^{-1} \left( \frac{1}{2} \right) - \left( \frac{1}{8} - \frac{1}{8} \tan^{-1}(2) + \frac{1}{2} \tan^{-1} \left( \frac{1}{2} \right) \right) \right) \\
&\quad + O \left( \frac{1}{N} \right) = \frac{\sigma^2}{2} + O \left( \frac{1}{N} \right) = \text{Var} \left( n_3^{(r,u)} \right) = \sigma^2 a^{(r,1)},
\end{aligned}$$

$$\begin{aligned}
\text{Var} \left( n_2^{(r,2)} \right) &= 2^{2j+2} \sigma^2 \int_{\frac{1}{2^{j+1}}}^{\frac{1}{2^j}} \int_0^{\frac{1}{2^{j+1}}} \frac{f_1^2}{f_1^2 + f_2^2} df_2 df_1 + O \left( \frac{1}{N} \right) = \sigma^2 \left( \frac{1}{2} + 2 \tan^{-1} \left( \frac{1}{2} \right) \right. \\
&\quad \left. - \frac{1}{2} \tan^{-1}(2) \right) + O \left( \frac{1}{N} \right) = \sigma^2 a^{(r,2)} + O \left( \frac{1}{N} \right) \approx 0.8737 \sigma^2.
\end{aligned}$$

Finally note that:

$$\begin{aligned}
\text{Var} \left( n_2^{(r,3)} \right) &= 2^{2j+2} \sigma^2 \int_0^{\frac{1}{2^{j+1}}} \int_{\frac{1}{2^{j+1}}}^{\frac{1}{2^j}} \frac{f_1^2}{f_1^2 + f_2^2} df_2 df_1 + O \left( \frac{1}{N} \right) = \sigma^2 \left( 1 - a^{(r,2)} \right) + O \left( \frac{1}{N} \right) \\
&\equiv \sigma^2 a^{(r,3)} \approx 0.1263 \sigma^2,
\end{aligned}$$

$$\text{Var} \left( n_{j,k_1,k_2}^{(r,1,4)} \right) = 2^{2j+2} \sigma^2 \int_0^{\frac{1}{2^{j+1}}} \int_0^{\frac{1}{2^{j+1}}} \frac{f_1^2}{f_1^2 + f_2^2} df_2 df_1 = \frac{\sigma^2}{2} + O \left( \frac{1}{N} \right) = \sigma^2 a^{(r,4)} + O \left( \frac{1}{N} \right).$$

Clearly we can find the variance of the second RT by permuting the order of the spatial variable in the integration, and this then completes the variance calculations. From the proofs of propositions (1) and (2) we can note that the components of  $\mathbf{n}^{(r,u)}$  for  $u = 1, 2, 3, 4$  are uncorrelated up to  $O \left( \frac{1}{N} \right)$ , and this can also be shown by direct calculation, *mutatis mutandis* the calculations given above. Thus as  $\epsilon_{\mathbf{x}}$  was zero-mean Gaussian, and we are forming linear combinations to obtain  $\mathbf{n}^{(r,u)}$ , the stated result follows from the expressions for the covariances of the components.

*Proof of Lemma 6*

a) First consider  $u = 1, 4$  so that the variance of the two Riesz components is  $1/2$ . Then by Lemma 5 it follows directly that  $(C_L^{(s)} + 1)M_{\xi}^{(\epsilon, r)2} \stackrel{\mathcal{L}}{=} \mathbf{Z}^{(r, u)T} \mathbf{Z}^{(r, u)} + O\left(\frac{1}{N}\right)$ ,  $T_1 = \mathbf{Z}^{(r, u)T} \mathbf{Z}^{(r, u)} \sim \chi_1^2 + \frac{1}{2}\chi_1^2 + \frac{1}{2}\chi_1^2$ .  $T_1$  has a Moment Generating Function (MGF) given by  $M_{T_1}(s) = \frac{1}{\sqrt{1-2s}} \frac{1}{1-s}$ , and thus  $f_{T_1}(t) = e^{-t} \frac{2}{\sqrt{\pi}} \int_0^{\sqrt{\frac{t}{2}}} e^{u^2} du$ . The probability that  $T_1$  does not exceed  $\lambda^2$  is given by

$$\begin{aligned} P(T_1 < \lambda^2) &= 1 + \left[ e^{-x} \int_0^{\sqrt{\frac{x}{2}}} \frac{2}{\sqrt{\pi}} e^{u^2} du \right]_{\lambda^2}^{\infty} - \int_{\lambda^2}^{\infty} \frac{1}{\sqrt{2\pi}} x^{-\frac{1}{2}} e^{-\frac{x}{2}} dx \\ &= 1 - \sqrt{\frac{8}{\pi}} \frac{1}{\lambda} e^{-\lambda^2/2} + e^{-\lambda^2/2} \left( O\left(\frac{1}{\lambda^3}\right) \right), \end{aligned} \quad (38)$$

where the cdf be found in Grad & Solomon [28][p. 472], and the function is expanded as  $\lambda \rightarrow \infty$ .

b) Consider now  $u = 2, 3$ . Wlog assume that  $a^{(r, u)} > 1/2$ , and otherwise relabel  $a^{(r, u)}$  and  $1 - a^{(r, u)}$  suitably. If  $a^{(r, u)} = 1 - a^{(r, u)}$  this collapses to the case given above. Define  $T' = T_1/2$  so that  $T' = \sum_{i=1}^3 a_i X_i^2$  where the  $X_i$  are iid Gaussian random variates with zero mean and unit variance, where  $\sum_{i=1}^3 a_i = 1$ . Then the MGF is by Grad & Solomon [28]

$$M_{T'}(s) = M_{T_1}(s/2) = \left(1 - 2(1 - a^{(r, u)})s/2\right)^{-\frac{1}{2}} \left(1 - 2a^{(r, u)}s/2\right)^{-\frac{1}{2}} (1 - 2s/2)^{-\frac{1}{2}} = \prod_{l=1}^3 (1 - 2a_l s)^{-\frac{1}{2}},$$

with  $a_1 = (1 - a^{(r, u)})/2$ ,  $a_2 = a^{(r, u)}/2$ ,  $a_3 = 1/2$ . Defining  $c_1 = 2/(1 - a^{(r, u)})$ ,  $c_2 = 2/a^{(r, u)}$ ,  $c_3 = 2$ , and hence as we assumed  $1 > a^{(r, u)} > 1 - a^{(r, u)}$  we may note that  $c_1 \geq c_2 \geq c_3$  and thus agrees with [28]'s notation. For future reference note that  $c_1 + c_2 = 2/(a^{(r, u)}(1 - a^{(r, u)}))$ ,  $u = 1, \dots, 4$ , and we may rewrite  $M_{T'}(s) = \prod_{j=1}^3 (1 - 2s/c_j)^{-1/2}$ . Using results from Grad and Solomon we may determine:

$$F_{T_1}(t) = F_{T'}(t/2) = \left(1 - Ae^{-c_3 t/4} \frac{1}{\sqrt{t}}\right) + o(1).$$

For suitably defined constant  $A$ . From these formulae we can thus consider the probability of an observation exceeding a large threshold, which will be necessary for the selection of an appropriate threshold.

*Proof of Lemma 7*

We establish

$$\text{Var}\left(n_2^{(h, 1)}\right) = \sigma^2 + O\left(\frac{1}{N}\right) = \text{Var}\left(n_l^{(h, u)}\right), \quad u = 1, 2, 3, 4, \quad l = 1, 2, 3, 4.$$

These results follow trivially from the form of the partial HT [21]. From the proofs of propositions (1) and (2) we can note that the components of  $\mathbf{n}^{(h, u)}$  for a fixed value of  $u = 1, 2, 3, 4$  are uncorrelated up to  $O\left(\frac{1}{N}\right)$ , and this can also be shown by direct calculation *mutatis mutandis* the calculations given above.

*Proof of Lemma 6*

$K$  wavelet coefficients will be thresholded where  $K_1$  magnitudes have the distribution given when  $u = 1, 4$  and  $K_2$  have the distribution that follows from  $u = 2, 3$ , where  $K_1 + K_2 = K$ , and  $K_1, K_2, K_3 = O(K)$ , so that by Dykstra [26]

$$P(\mathcal{M}_K < \lambda_K^2) = \left(1 - \sqrt{\frac{8}{\pi}} \frac{1}{\lambda_K} e^{-\lambda_K^2/2}\right)^{K_1} \left(1 - Ae^{-c_3 \lambda_K^2/4} \frac{1}{\lambda_K}\right)^{K_2} = \left[1 - \frac{A_2}{\lambda_K} e^{-\lambda_K^2/2}\right]^{K_3},$$

ignoring  $o(1)$  terms in  $K$  for suitably chosen constant  $A_2$ , if  $\lambda_K = O(\log[K])$ . Thus with  $\lambda_K^2 = 2 \log[K] + C_2 \log[\log[K]]$ ,

$$\begin{aligned} P(\mathcal{M}_K < \lambda_K^2) &= \left(1 - A_2 e^{-\lambda_K^2/2} \frac{1}{\lambda_K}\right)^{K_3} + o(1) \\ &= \left(1 - A_2 e^{-(2 \log[K] + C_2 \log[\log[K]])/2} \frac{1}{\sqrt{2 \log[K] + C_2 \log[\log[K]]}}\right)^{K_3} \rightarrow 1, \end{aligned}$$

if  $C_2/2 > -1/2$ , i.e.  $C_2 > -1$ , and we take  $C_2 = 0$ .

#### D: RISK CALCULATIONS

##### *Proof of Theorem 1*

We firstly note from [29][p. 293] that the risk of regular hard thresholding is given by

$$\begin{aligned} R_\theta^{(c)}(\lambda) &= \int_{(w+\theta_1)^2 \geq \lambda^2} w^2 \phi(w) dw + \theta_1^2 \int_{(w+\theta_1)^2 < \lambda^2} \phi(w) dw \\ &= 1 + \int_{(w+\theta_1)^2 < \lambda^2} [\theta_1^2 - w^2] \phi(w) dw. \end{aligned} \quad (39)$$

We may then note that the risk for the hyperanalytic threshold with  $\text{var} \left( n_l^{(s,u)} \right) = \sigma^2 \sigma_l^2$ , for  $l = 0, \dots, L$ , where  $\sigma_l^2$  takes the value 1 or  $a^{(s,u)}$ , is given by:

$$\begin{aligned} R_\theta^{(s)}(\lambda) &= \theta_1^2 \int_{\sum_l w_l^2 \leq \lambda^2} \prod_l \sigma_l^{-1} \phi \left( \frac{w_l - \theta_l}{\sigma_l} \right) dw_l + \int_{\sum_l w_l^2 > \lambda^2} [w_1 - \theta_1]^2 \prod_l \sigma_l^{-1} \phi \left( \frac{w_l - \theta_l}{\sigma_l} \right) dw_l \\ &= 1 + \int_{\sum_l (w_l + \theta_l)^2 \leq \lambda^2} [\theta_1^2 - w_1^2] \prod_l \sigma_l^{-1} \phi \left( \frac{w_l}{\sigma_l} \right) dw_l. \end{aligned} \quad (40)$$

##### *Proof of Corollary 1*

$$R_0^{(c)}(\lambda) = 1 - \int_{-\lambda}^{\lambda} w^2 \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}w^2} dw = 1 - \gamma \left( 1/2, \frac{1}{2}\lambda^2 \right) = e^{-\frac{1}{2}\lambda^2} \left( \frac{\sqrt{2}}{\sqrt{\pi}\lambda} + O(\lambda^{-3}) \right). \quad (41)$$

$$R_0^{(a)}(\lambda) = 1 - \int_{w_1^2 + w_2^2 \leq \lambda^2} w_1^2 \frac{1}{2\pi} e^{-\frac{1}{2}(w_1 + w_2)^2} dw_1 dw_2 = e^{-\frac{1}{2}\lambda^2} \left( 1 + \frac{1}{2}\lambda^2 \right).$$

Furthermore the risk at  $\theta = \mathbf{0}$  can also be found for the other hyperanalytic thresholds. We note that for  $s = r$  and  $u = 1, 4$ , denoted by  $s = r_1$  in the Figures:

$$\begin{aligned} R_0^{(r_1)}(\lambda) &= 1 - \int_{w_1^2 + w_2^2 + w_3^2 \leq \lambda^2} w_1^2 \frac{2}{(\sqrt{2\pi})^3} e^{-\frac{1}{2}(w_1^2 + 2(w_2^2 + w_3^2))} dw_1 dw_2 dw_3 \\ &= 1 - \int_{-\lambda}^{\lambda} w_1^2 \frac{2}{(\sqrt{2\pi})^3} e^{-\frac{1}{2}w_1^2} \left[ \int \int_{w_2^2 + w_3^2 \leq \lambda^2 - w_1^2} e^{-(w_2^2 + w_3^2)} dw_2 dw_3 \right] dw_1 \\ &= 1 + \gamma \left( 1/2, \frac{1}{2}\lambda^2 \right) + 4\lambda^2 \left( \Phi(\sqrt{2}\lambda) - \Phi(\lambda) \right) - 2\gamma(1/2, \lambda^2) = e^{-\frac{1}{2}\lambda^2} \left( \frac{4\sqrt{2}}{\sqrt{\pi}\lambda} + O(\lambda^{-3}) \right), \end{aligned} \quad (42)$$

whilst for  $u = 2, 3$  denoted by  $s = r_2$ ,

$$R_0^{(r_2)}(\lambda) = 1 - \int_{\sum_l w_l^2 \leq \lambda^2} w_1^2 \frac{1}{(\sqrt{2\pi})^3 \sqrt{a^{(r,u)}(1 - a^{(r,u)})}} e^{-\frac{1}{2}(w_1^2 + w_2^2/a^{(r,u)} + w_3^2/(1 - a^{(r,u)}))} dw_1 dw_2 dw_3$$

$$\begin{aligned} R_0^{(h)}(\lambda) &= 1 - \int_{w_1^2 + w_2^2 + w_3^2 + w_4^2 \leq \lambda^2} w_1^2 \frac{1}{(\sqrt{2\pi})^4} e^{-\frac{1}{2}\sum w_i^2} dw \\ &= 1 - \int_0^\lambda \int_0^{2\pi} \int_0^\pi \int_0^\pi r^2 \cos^2(\theta) \frac{1}{(\sqrt{2\pi})^4} e^{-\frac{1}{2}r^2} r^3 \sin^2(\theta) \sin(\phi) dr d\theta d\phi d\lambda \\ &= 1 - \frac{(\frac{\pi}{4})(2\pi)}{4\pi^2} \int_0^\lambda r^5 e^{-\frac{1}{2}r^2} dr = 1 - \frac{1}{8} \int_0^{\frac{1}{2}\lambda^2} 4s^2 e^{-s} ds = e^{-\frac{1}{2}\lambda^2} \left( 1 + \frac{1}{2}\lambda^2 + \frac{1}{8}\lambda^4 \right). \end{aligned} \quad (43)$$